DOCUMENT RESUME

ED 174 976 CS 004 986

AUTHOR Harper, Betty
TITLE The Construction of an Instrument to Measure Elementary Teachers' Knowledge of a Subskills Approach to Teaching Reading. Technical Report No. 501.
INSTITUTION Wisconsin Univ., Madison. Research and Development Center for Individualized Schooling.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Aug 79
GRANT OB-NIE-G-78-0217
NOTE 189p.

EDRS PRICE MF01/PC08 Plus Postage.
DESCRIPTORS Elementary Education; *Elementary School Teachers; *Measurement Techniques; Reading Skills; *Reading Tests; Research Methodology; Teacher Improvement; Teacher Qualifications; *Teaching Skills; *Test Construction; *Test Validity

ABSTRACT

An instrument to assess elementary school teachers' knowledge of a subskills approach to teaching reading has been developed. In the first phase of the development, an outline of test content was devised and submitted to four reading authorities for evaluation. The second phase involved writing and selecting items for the preliminary test form. Based on the test outline, a pool of multiple choice items was generated and subsequently reviewed by persons knowledgeable in the subject matter area. The items were then revised or replaced based on the reviewers' comments. A further check on content validity was made when knowledgeable persons keyed the items and furnished information for improving the preliminary test form. The third phase involved field testing the preliminary test form with elementary teachers, analyzing and revising the items based on the field tryout data, and reassessing content validity of the revised instrument. In the final phase, the revised instrument was subjected to a field tryout with 104 teachers. The resulting data were examined to determine reliability and content validity. A review of the test development process indicated that it was highly probable that the instrument does demonstrate content validity. This conclusion is based on the cumulative evidence that content validity was built into the instrument from the beginning. (Author/FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED174976

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Technical Report No. 501

THE CONSTRUCTION OF AN INSTRUMENT TO MEASURE ELEMENTARY TEACHERS' KNOWLEDGE OF A SUBSKILLS APPROACH TO TEACHING READING

by

Betty Harper

Report from the Project on
Studies in Language: Reading and Communication

Wayne Otto
Faculty Associate

CS504986

Wisconsin Research and Development Center
for Individualized Schooling
The University of Wisconsin
Madison, Wisconsin

August 1979

This Technical Report is a doctoral dissertation reporting research supported by the Wisconsin Research and Development Center for Individualized Schooling. Since it has been approved by a University Examining Committee, it has not been reviewed by the Center. It is published by the Center as a record of some of the Center's activities and as a service to the student. The bound original is in the University of Wisconsin Memorial Library.

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Center Grant No. OB-NIZ-G-78-0217

MISSION STATEMENT

The mission of the Wisconsin Research and Development Center is to improve the quality of education by addressing the full range of issues and problems related to individualized schooling. Teaching, learning, and the problems of individualization are given concurrent attention in the Center's efforts to discover processes and develop strategies and materials for use in the schools. The Center pursues its mission by

- conducting and synthesizing research to clarify the processes of school-age children's learning and development
- conducting and synthesizing research to clarify effective approaches to teaching students basic skills and concepts
- developing and demonstrating improved instructional strategies, processes, and materials for students, teachers, and school administrators
- providing assistance to educators which helps transfer the outcomes of research and development to improved practice in local schools and teacher education institutions

The Wisconsin Research and Development Center is supported with funds from the National Institute of Education and the University of Wisconsin.

WISCONSIN RESEARCH AND DEVELOPMENT
CENTER FOR INDIVIDUALIZED SCHOOLING

ACKNOWLEDGEMENTS

The author wishes to express sincere gratitude to Dr. Wayne Otto, chairman of the dissertation committee, for his guidance and assistance in the completion of this dissertation. His unselfish contribution of time and expertise in guiding the author to acquire the knowledge and skills necessary to complete this project are deeply appreciated. Appreciation is also expressed to the other committee members: Dr. Dale D. Johnson, Dr. Penelope Peterson, Dr. Sarah Tarver, and Dr. Kenneth Dulin for their guidance and encouragement.

Special thanks is given to the administrators and teachers of the participating school districts who cooperated in the field tryouts of the instrument.

Finally, the author is especially indebted to those persons who generously gave of their time and talents in reviewing the test items and in offering suggestions for their improvement. The cooperation of Karlyn Kamm, Sandra White, Roger Eldridge, and Carol Dana in reviewing the items and Petey Young in editing the items is sincerely appreciated. To the persons who further assessed content validity of the items and then acted in a consultant role, the author is genuinely grateful.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	iv
LIST OF TABLES	ix
ABSTRACT	xi
CHAPTER	
I INTRODUCTION	1
Review of the Literature	3
Review of Available Instruments	3
Measures of Word Identification and Other Reading Skills	4
Measures of Diagnostic Proficiency	9
Global Measures of the Content of Reading Instruction	11
Critique of Existing Instruments	12
Justification for Tapping Knowledge of a Subskills Approach	25
1. Reading is a complex skill which is composed of specific subskills	26
2. A skill-hierarchy model provides a valid tool for viewing reading as a system of interrelated and sequenced components	29
3. A skill-centered approach provides a framework for individualizing instruction	33
4. A skill-centered approach contributes to general reading performance	35

TABLE OF CONTENTS
(Continued)

CHAPTER		Page
I	(Continued)	
	5. Direct explicit instruction, which is inherent in a subskills approach, facilitates learning to read	38
	Summary.	41
	Statement of the Problem	43
II	INSTRUMENT DEVELOPMENT.	46
	Developing the Test Specifications	46
	Purpose of the Test	47
	Test Outline.	47
	Item Format	57
	Number of Items	58
	Item Evaluation	59
	Constructing and Selecting Items	60
	Identifying Topics and Content.	60
	Writing the Items	62
	Reviewing and Revising the Items. . . .	64
	Field Tryout I, Analysis, and Revision of the Preliminary Test Form	68
	Examinees	68
	Administration.	69

TABLE OF CONTENTS
(Continued)

CHAPTER		Page
II	(Continued)	
	Analysis of the Data.	70
	Revising the Preliminary Test Form. . . .	77
	Reassessing Content Validity.	78
	Field Tryout II.	78
	Examinees	78
	Administration.	82
	Analysis of Data.	82
III	PRESENTATION, ANALYSIS AND DISCUSSION OF DATA . . .	84
	Evidence of Reliability.	84
	Evidence of Validity	92
	Demographic Data	95
	Summary.	100
IV	SUMMARY, CONCLUSIONS, LIMITATIONS, AND RECOMMENDATIONS	102
	Summary of the Research.	102
	Conclusions.	104
	Limitations.	106
	Recommendations.	106
REFERENCES		112

TABLE OF CONTENTS
(Continued)

APPENDICES		Page
A	EVALUATION FORM FOR TEST OUTLINE.	121
B	SOURCE MATERIALS FOR ITEM CONTENT AND EVALUATION FORM	124
C	INSTRUCTIONS AND EVALUATION FORM FOR ITEM REVIEW	126
D	ITEMS FOR EACH CONTENT AREA IN THE PRELIMINARY TEST FORM	130
E	TEST FORM USED FOR FIELD TRYOUT I	132
F	INSTRUCTIONS FOR EXAMINERS -- FIELD TRYOUT I. . .	153
G	ITEMS FOR EACH CONTENT AREA IN THE REVISED TEST FORM	156
H	INSTRUCTIONS AND EVALUATION FORM FOR REASSESSMENT OF CONTENT VALIDITY.	158
I	TEST FORM USED FOR FIELD TRYOUT II.	160
J	INSTRUCTIONS FOR EXAMINERS -- FIELD TRYOUT II . .	175

LIST OF TABLES

TABLE		Page
1	Teacher Knowledge Tests	13
2	Difficulty Levels on a Biserial Correlation Coefficient Values for Items in Field Tryout I. . . .	71
3	Items Grouped by Biserial Correlation Coefficients and Difficulty Levels -- Field Tryout I.	76
4	Number of Reading Courses Reported by Second Field Tryout Sample	80
5	Years of Elementary Teaching Experience of Second Field Tryout Sample.	81
6	Difficulty Levels and Biserial Correlation Coefficients of Items in Field Tryout II.	85
7	Items Grouped by Biserial Correlation Coefficients and Difficulty Levels -- Field Tryout II	90
8	Descriptions on Variables of Experience and Number of Reading Courses for Samples in Field Tryouts of the Test	96
9	Comparison of the Difference Between Sub-Group Mean Scores for Number of Reading Courses of Second Tryout Sample.	98

THE CONSTRUCTION OF AN INSTRUMENT TO MEASURE

ELEMENTARY TEACHERS' KNOWLEDGE

OF A SUBSKILLS APPROACH

TO TEACHING READING

ABSTRACT

This is a report of the procedures followed in constructing an instrument to assess elementary teachers' knowledge of a subskills approach to teaching reading. In the first phase of instrument development an outline of test content was developed, and was submitted to four reading authorities for evaluation. The second phase involved writing and selecting items for the preliminary test form. Based on the test outline, a pool of multiple choice items was generated and subsequently reviewed by persons knowledgeable in the subject matter area. The items were then revised and/or replaced, based on the reviewers' comments. A further check on content validity was made when knowledgeable persons keyed the items and furnished information for improving the preliminary test form. The third phase of the test construction process involved: (1) field testing the preliminary test form with elementary teachers, (2) analyzing and revising the items based on the field tryout data, and (3) reassessing content validity of the revised

instrument. In the final phase of test construction the revised instrument was subjected to a field tryout with 104 elementary teachers. The resulting data were examined to determine reliability and content validity of the present instrument. In addition, a one-tail t-test was used to determine whether differences between the group mean scores for teachers with zero or one reading course and teachers with three or more reading courses were significant at the .05 level. The resulting separate variance estimate indicated a t-value of -.31, less than the critical value of t (1.98) required for significance at the .05 level. The correlations between number of reading courses and years of teaching experience to test scores were also examined. The Pearson product-moment correlation between number of reading courses and performance on the instrument was computed at .14 while the correlation between years of teaching experience and performance on the instrument was -.31. An examination of data from the second field tryout indicated that the present instrument, with a Hoyt internal consistency reliability coefficient of .82 and a standard error of measurement of 3.29, may be considered adequate for interpreting individual measurement. Further, a review of the test development process indicated that it is highly probable that the present instrument demonstrates content validity. This conclusion is based

or the cumulative evidence that content validity was built into the instrument from the beginning of the test development process by systematic, logical analysis for determining the adequacy of the items for measuring content in the test outline.

Wayne Otto

Wayne Otto, Major Professor

CHAPTER I

INTRODUCTION

Reading educators generally assume that one basis for effective reading instruction is knowledge about what it is that one is teaching (Durkin, 1976; Artley and Hardin, 1975). The implication is that lack of knowledge results in inferior instruction. The importance attached to acquiring a knowledge competency in basic reading instruction is evident in the trend to require more reading courses in pre-service training programs and in teacher certification regulations (Morrison and Austin, 1976; Rafferty, 1975). And, related to this trend, the need to assess teachers' knowledge has led to the development of instruments designed to assist in that process.

Although a number of instruments have been developed to quantify teachers' knowledge, at the present time we do not know what teachers know about the teaching of reading. This is because the tests that have been constructed are unfocused and, consequently, the content of reading instruction has not been systematically examined. An examination of the available tests reveals the confusion about what constitutes "knowledge competency." A common core of knowledge is not emphasized, but rather, instruments have been developed for measuring teachers' knowledge of word identification skills, diagnostic procedures, and general content of reading instruction. While a majority of the instruments assess word identification skills and one might assume that they test common content, the components

included in the test base vary from test to test. Consequently, because there is little agreement as to the knowledge base teachers should possess, whether or not a teacher is "knowledgeable" depends on what instrument is used.

Coupled with the lack of consensus concerning what content of reading instruction is most important is the absence of instruments for assessing knowledge of how to teach reading. Test authors assume that if teachers know content they have acquired the basis for effective instruction. Because knowledge of the process of teaching reading is not tapped, it appears that one cannot assume that teachers have the basis for a teaching competency.

While one could advocate either a wholistic or a subskills approach to teaching reading, knowledge of a subskills approach seems more important for a number of reasons. First, the importance of subskills in the developmental sequence of skill attainment has historical, consensual, and empirical support. Essentially, before one deals with wholes smaller aspects have to be mastered first. Second, a subskills approach allows one to proceed toward the goal of reading maturity in a rational manner. Instruction is individualized as a student learns the skills that are required for competent reading at his/her own rate. Finally, direct systematic instruction provides the means for moving individual students to independence in their reading comprehension.

Thus, while the improvement of pre-service teacher education has led to the specification and measurement of teachers' "knowledge

competency", findings about teachers' knowledge are inconclusive and piecemeal because of the particular instruments that have been constructed. In the section that follows, a critical review of the available knowledge tests will point out their lack of focus and their overconcern with what to teach. Then, support for a knowledge base that focuses on the teaching of reading, specifically a sub-skills approach, will be provided.

Review of the Literature

The review of literature is divided into three sections. The first section is concerned with the description of the 18 instruments that have been constructed to essentially measure elementary teachers' knowledge of the content of reading instruction. The second section deals with a critical analysis of the available instruments. Finally, in the third section, a review of selected literature provides justification for tapping knowledge of a skill-centered approach to teaching reading.

Review of Available Instruments

This section provides a review of the instruments that have been constructed to essentially measure elementary teachers' knowledge of the content of reading instruction. Because these instruments have been designed to focus on three specific areas, they will be presented in three sections: (1) measures of word identification and other reading skills, (2) measures of diagnostic proficiency,

and (3) global measures of the content of reading instruction.

Measures of Word Identification and Other Reading Skills

Teachers' knowledge of word identification skills was investigated in at least 10 studies while knowledge of word identification skills and other reading skills was assessed in two other studies.

Word Identification Skills. Shannon (1959) constructed three tests of 50 four-option, multiple choice items. One hundred items measured visual-aural discrimination of phonemes in words and 50 items tested ability to select phonetic respellings of key nonsense words. According to Shannon, the phonic principles that were tested were those included in some texts on the teaching of reading and were common to basal reading series. Based on a split-half technique, the reliability coefficient and standard error of measurement, respectively, for each test was: test one, .90 and 2.67; test two, .91 and 5.65; test three, .88 and 2.96. The sample consisted of 449 college students majoring in elementary education and 40 in-service teachers.

Shubert (1959) devised an informal 10-item quiz and administered it to a sample of 80 elementary teachers and 42 secondary teachers. Examinees were expected to name the vowels, list several words with short and long vowels, define the term long vowel, and state any syllabication principles they might know.

Aaron (1960) constructed 60 multiple choice items involving nonsense words to assess knowledge of six vowel principles and two

consonant principles taught in basal readers in grades two and three. The sample of 293 examinees included pre-service and in-service elementary teachers, secondary teachers, and administrators.

Farinella (1960) constructed 84 selection-type and supply-type items[1] to measure knowledge of phonic analysis and structural analysis skills that were common to four basal reading series. A sample group of 394 teachers in grades one through six were required to define terms, to state generalizations and to apply that knowledge. Farinella reported a reliability coefficient of .91, based on the split-half technique.

Based on several basal reader manuals, Ramsey (1963) constructed an 85-item instrument which was administered to 236 elementary education students. Thirty items tested knowledge of symbol-sound relationships with examinees spelling unfamiliar words pronounced by the examiner while 55 multiple choice items tested application of phonic and syllabication principles as well as ability to select phonic terms and phonic rules.

Durkin (1964) constructed a test of 10 brief subtests that was administered to 603 students enrolled in reading methods courses. About two-thirds of the items in Durkin's instrument tested knowledge of vowel principles, one-third tested knowledge of consonant principles, and four items tested knowledge of four different

---

[1]Selection-type items are represented by the true-false, multiple choice, and matching forms. Supply-type items are represented by the short-answer form.

syllabication principles. Included in the test are rather esoteric elements such as the two sounds represented by qu and the three sounds represented by x, while knowledge of more utilitarian elements (e.g., r principles and the schwa sound) are not assessed. In addition, 39 of the 60 responses require underlining or checking correct answers while the balance are open-ended.

McCollum (1964) put together a battery of five tests with a total of 140 multiple choice items which was administered to 230 pre-service and in-service elementary teachers. McCollum states that in the first 30-item subtest the examinee reacts to 20 initial consonant blends, two triple consonant blends, three digraphs, four consonant rules, and 14 vowel rules. The second subtest assesses ability to use a dictionary pronunciation key, and the third subtest assesses ability to identify correctly syllabicated nonsense words. The fourth subtest requires examinees to identify root words and the final subtest assesses knowledge of linguistic terminology. According to the author, the syllabication content is common to three basal reader series while the linguistic terms are common to literature in that area. McCollum applied the Kuder-Richardson formula to the results of the experienced teachers and obtained a reliability coefficient of .90.

Spache and Baggett (1965) modified Aaron's phonics test to 50 items, constructed a 65-item test in which examinees wrote the number of syllables present in words, and devised a test of 12 multiple choice items on syllabication rules. The instrument was administered

to a sample of 99 in-service teachers and graduate students.

Henriksen (1968) developed a 170-item battery of eight subtests that was administered to a sample of 179 elementary and secondary teachers and administrators. The tests were based on word recognition skills listed in a guide published by the Georgia State Department of Education. The test base includes word recognition terminology, root words, compound words, suffixes, syllables, context clues, application of phonics knowledge, and dictionary use. Item forms are the selection-type and the supply-type. Test reliability, based on a split-half technique, was .95.

Mazurkiewicz (1975) developed two instruments in which selection-type and supply-type items were used to measure knowledge of phonics. The first instrument was a seven-item questionnaire to measure knowledge of the number of vowel and consonant letters and sounds and three frequently used phonic generalizations. The second instrument was a 45-item phonic test in which examinees had to justify 27 of their answers. The phonic test includes items on the four sounds of s, three sounds of ch, and the four spellings of the /sh/ sound and the /zh/ sound. The sample consisted of 252 teachers in kindergarten through sixth grade.

Johnson (1976) developed an informal screening device of 49 selection-type and supply-type items to assess teachers' knowledge of word identification skills. Twenty-six items assess ability to discriminate vowel phonemes in synthetic words. The remaining items assess skill in phonic syllabication, structural analysis, and

contextual analysis. This instrument was not field tested.

Additional reading skills. Finally, two instruments were developed to investigate knowledge of word identification skills and other reading skills.

Broman (1962) based her test of 34 selection-type items on a common core of skills in 10 basal reader manuals. Examinees were required to select words to define phonic and syllabication principles and terms. In addition, 20 items dealt with knowledge of how skills are sequenced in basal reader texts. The examinee group consisted of 113 teachers in grades one through six.

Thompson (1972) constructed a test of 100 four-choice, multiple choice items which was administered to 171 teachers of early childhood through college levels. It was based on the Florida Catalog of Reading Objectives which is used to assess reading skills in grades kindergarten through twelve. Application of the following skills is tested: decoding, word processing, reading comprehension, and work study. A disproportionately large number of items (40) are devoted to decoding. Although each of the other areas is broken into a number of component skills the usual practice is to allot one item to measure a particular skill. Fifty-five of the 100 items fell outside a .40-.90 difficulty range and had discrimination indices below .30. Thompson reported a Kuder-Richardson Formula 20 reliability coefficient of .91.

Measures of Diagnostic Proficiency

Three instruments were developed to measure the diagnostic proficiency of elementary school teachers. In examining these instruments one notes a lack of agreement as to a common test base. In addition, the tests differ in the emphasis that is placed on formal and informal assessments as well as the range of diagnostic assessments chosen for inclusion.

Wade (1960) constructed 10 subtests varying in task orientation to measure a teacher's problem-solving ability. A combined sample of 176 teachers and 186 undergraduates with and without teaching experience determined if texts are of appropriate difficulty, ranked selected passages in order of increasing difficulty, and formed three reading groups on the basis of oral reading accuracy. Teachers also judged reading growth, answered questions about oral reading behavior, selected word identification skills and rules a child should learn, categorized oral reading errors on sight words, and matched skill development exercises to work recognition errors. The 73 items are of the selection-type in which teachers rank order four choices. Based on a split-half technique, test reliability was .87 for the total sample of three subgroups (practice teachers, undergraduates, and teachers in grades one through five).

Burnett's (1961) instrument measures a teacher's ability to utilize diagnostic procedures to solve interrelated problems. The first set of problems deals with word analysis difficulties of a third grade student and the second set of problems deals with

personality and comprehension weaknesses of a fifth grade student. Examinees are required to: (1) select useful information from a pool of data, (2) obtain additional information based on the initial data, (3) interpret these data, (4) make a recommendation on the basis of these data, and (5) reevaluate recommendations given additional data. Item forms include both the selection-type and the supply-type. The test was administered to 187 examinees. Test reliability, based on a split-half technique, was .76 for the sample of experienced teachers, .84 for the reading specialists, and .33 for undergraduate students.

Thomas (1975) constructed a 70-item instrument which requires examinees to choose data sources and to interpret test data for four assessments in reading. These include determination of: (1) reading levels and basic reading groups, (2) progress, (3) readiness, and (4) comparisons between reading expectancy levels and general reading achievement. The test is lengthy with multiple choice items requiring from one to four answers. Items are complicated and often examinees are required to consult footnotes and data on several pages in order to answer items. The sample consisted of 68 elementary teachers, reading teachers, graduate students, and "others". Thomas reported a Kuder-Richardson Formula 20 reliability coefficient of .84 and a standard error of measurement of 3.77. She also reported a Livingston criterion-referenced reliability coefficient of .91.

Global Measures of the Content of Reading Instruction

Mallett (1972) constructed a 20-item test which was administered to 22 first grade teachers to determine the extent of their knowledge of reading instruction and methodology. The items tap knowledge of word identification and comprehension skills as well as general information about reading. In this author's opinion, the test items are not representative of the content of reading instruction that is most valid for first grade teachers to know. Mallett indicates that teacher knowledge as measured by this test was not significantly related to progress in reading for the first 18-week period of the first grade experience.

Artley and Hardin (1975) developed a 97-item test of seven content areas identified in reading methods texts:

-- the reading act
-- preparing for reading
-- word identification
-- comprehension and critical reading
-- study skills and reading in the content areas
-- development of interests and tastes
-- overcoming difficulties

Since the authors do not identify items with content areas the instrument has little diagnostic value because so little information is provided by the global test score. In addition, the relevance or appropriateness of the items for assessing each of the content areas cannot be determined because the authors fail to define the

content areas. The instrument was administered to 552 examinees. A Kuder-Richardson Formula 20 reliability coefficient of .92 was reported for the total sample of three subgroups: in-service elementary teachers, elementary teachers taking graduate courses, and elementary education seniors. Finally, Kingston, Brosier and Hsu (1975) concluded that because factor analysis failed to identify the seven areas they either "are not bona fide factors in the sense of being unitary characteristics or the instrument itself fails to measure them accurately" (p. 136).

Data for the instruments reviewed are summarized in Table 1.

Critique of Existing Instruments

Although the importance of acquiring a knowledge base for teaching reading has been acknowledged as important there seems to be little agreement as to the knowledge base. An examination of the existing tests indicates that knowledge of word identification skills, diagnostic procedures, or general content of reading instruction have all been proposed as the basis for effective instruction. Knowledge of word identification skills has been investigated in 12 of the 18 instruments, which indicates a tendency to agree that this component is important. But, an analysis of this test base in individual instruments reveals little agreement as to what aspects of this component are most important. Some instruments assess knowledge of phonic generalizations while other tests may assess knowledge of both phonics and syllabication principles or may

Table 1

Teacher Knowledge Tests

Author	Year	No. Items	Content	Type of Items/Task
Shannon	1959	150	Word Identification: Phonics	Selection: Multiple Choice
Shubert	1959	10	Word Identification: Phonics Syllabication	Supply: Word(s); State Rule
Aaron	1960	60	Phonics	Selection: Multiple Choice
Farinella	1960	84	Word Identification: Phonics Structural Analysis Syllabication	Selection: Multiple Choice Supply: Short Answer State Rule
Burnett	1961	10 problems (40 items)	Diagnostic Proficiency: Select & Interpret Data Make Recommendations	Selection: Rank order 4 responses

Table 1 (Continued)

Author	Year	No. Items	Content	Type of Items/Task
de	1960	10 problems (73 items)	Diagnostic Proficiency: Ten Skills (P. 9)	Selection: Choose 1 of 3-5 Responses; Rank Order 5 Responses; True-False; Check Correct Answer; Matching; Classify Supply: Word; State Rule
oman	1962	34	Word Identification: Phonics Syllabication Sequencing Skills in Basal Readers	Selection: Multiple Choice; Matching; Rank Order (3); Check Correct Answer; Choose Answer for Blank
msey	1963	85	Word Identification: Phonics Structural Analysis	Selection: Multiple Choice Supply: Spell Unfamiliar Syllables
Collum	1964	140	Word Identification: Phonics Structural Analysis Syllabication Dictionary Skills Linguistic Terms	Selection: 1 of 3-4 Options Multiple Choice Check Correct Answer

Table 1 (Continued)

Author	Year	No. Items	Content	Type of Items/Task
Durkin	1964	60	Word Identification: Phonics Syllabication	Selection: Underline Answer Check Answer Identify Answer With 0/+ Supply: Syllable Division Word; State Rule
Spache and Baggett	1965	127	Word Identification: Phonics Syllabication	Selection: Multiple Choice Supply: Number of Syllables
Henriksen	1968	170	Word Identification: Phonics Structural Analysis Contextual Analysis Syllabication Dictionary Skills	Selection: Multiple Choice Supply: Word Part
Mallet	1972	20	Word Identification: Phonics Contextual Analysis Syllabication Comprehension General Content	Selection: Multiple Choice

Table 1 (Continued)

Author	Year	No. Items	Content	Type of Items/Task
hompson	1972	100	Word Identification: Phonics Structural Analysis Contextual Analysis Dictionary Skills Comprehension Work Study Skills	Selection: Multiple Choice
azurkiewicz	1975	52	Phonics	Selection: Multiple Choice Supply: Number Reason for Answer Choice
homas	1975	70	Diagnostic Proficiency: Choose Data Source Interpret Test Data	Selection: Multiple Choice (Items Require 1 to 4 Answers)
rtley & ardin	1975	97	Global Reading Knowledge: Seven Areas (P.11)	Selection: Multiple Choice
ohnson	1976	49	Word Identification: Phonics Structural Analysis Contextual Analysis Syllabication	Selection: 1 of 3-4 Options Supply: Word in a Sentence State Rule Syllable Division

add various combinations of word identification skills to the test base. Finally, even though different aspects of the word identification component are investigated in various tests, knowledge of phonics is assessed in each of the instruments. This focus on phonics content can be attributed, in part, to the controversy in the late 50's and 60's that centered on the who, what, when, where, and why of phonics instruction. With this emphasis on the importance of phonic analysis for learning to read, phonics content was given more coverage in reading methods courses and instruments were constructed to measure the extent of teachers' knowledge of phonics and other word identification skills. However, inclusion of phonics content does not necessarily predicate agreement as to the generalizations that are most important for teachers to know. The generalizations that are emphasized vary from test to test, with the basis for choosing them appearing to be the subjective judgment of the test author rather than based on a measure of their utility in the classroom. Knowledge of some of the least useful generalizations (Mazurkiewicz, 1975) (e.g., the y at the end makes the preceding vowel long, if two vowels are adjacent in a word the first vowel is long) are assessed by Farinella (1960), Aaron (1960), Durkin (1964), Spache and Baggett (1965), and Henriksen (1968). Rather esoteric elements are investigated by several authors. For example, in her test Durkin (1964) included the two sounds of qu and three sounds of x while failing to assess knowledge of the schwa sound and the r principle. Mazurkiewicz (1975) assessed knowledge of the four

sounds of s, three sounds of ch, and four spellings of the /sh/ sound and the /zh/ sound. He also determined if teachers were aware of the number of vowel sounds (11 to 24) and the number of consonant sounds (23 or 24). Thompson (1975) tested knowledge of the /ch/ sound spelled either tch, ti, or t, /u/ sound spelled ou as in would, and the schwa sound spelled o as in lion.

Even though it would seem reasonable that generalizations that are most useful as an aid to word identification (Clymer, 1963; Emans, 1967; Bailey, 1967) should form a common core in these tests one explanation for the tendency to include inadequate generalizations can be ascribed to the test authors' reliance on common skill content in basal reader texts as the determiner of the test base. This appears to be a poor practice because it introduces the bias of the basal text authors in regard to the generalizations they chose to emphasize in the series. Some generalizations that are very popular with basal text authors are very ineffectual (Spache & Spache, 1977, p. 380) and their inclusion in a knowledge test is not warranted. Also, basal text authors may state generalizations in gross terms or may break them down into finer terms and state a large number of exceptions which further alters a test base. In addition to faulting test content that is based on skills common to basal series, a test base derived from outdated texts may not be particularly relevant. Because Mazurkiewicz (1975) judges the adequacy of responses on the basis of definitions found in a 1950 reading text and two dictionaries dated 1964 and 1967, the relevance

of the test content is questioned. Thus, the practice of utilizing basal reader texts or outdated texts is not only questionable but may not provide a valid indicator of examinees' knowledge or of what it is important for them to know. The argument can then be raised that examinees' scores may vary considerably, considering the particular source materials that the test constructor consults.

The lack of agreement as to a common test base can also be noted in the three measures of diagnostic proficiency (Wade, 1960; Burnett, 1961; Thomas, 1975). These tests differ in the emphasis that is given to informal and formal procedures as well as the range of diagnostic assessments that are made. Wade (1960) did not test teachers' proficiency in interpreting formal assessment measures but focused instead on analyzing oral reading behavior of individual students. Burnett (1961) initially emphasized interpretation of data from standardized tests for analyzing two students' reading problems while subsequent data were obtained from informal sources. On the other hand, Thomas (1975) primarily emphasized interpretation of data obtained from formal sources and dealt with assessment in a broader sense than either Wade or Burnett. These assessments ranged from the determination of reading readiness, reading levels and grouping to comparisons between reading expectancy levels and reading achievement.

A second criticism of the available instruments focuses on the types of tasks used by the test author. While test items are either the selection-type or the supply-type, tasks that are used

to measure a teacher's knowledge competency seem to vary as much as the content to be tested. Supply-type items may require examinees to provide a word part, single word, word in a sentence, symbol (e.g., syllable or accent mark), or an answer and statement of a rule or reason. Items patterned after the selection-type may require examinees to choose one or more options in a multiple choice format, rank order four or five responses, check, circle, or underline an answer, identify answers with + or 0, match options, choose an answer for a blank, classify, or designate true or false. While three types of supply tasks were the most included in an instrument (Durkin, 1964; Johnson, 1976), four authors included three to seven selection-type tasks (Wade, 1960; Broman, 1962; McCollum, 1964; Durkin, 1964). Of the eight authors that included both selection and supply items in a test, Johnson (1976) provided four different kinds of tasks, Durkin (1964) provided six, and Wade (1960) provided eight.

Requesting teachers to respond to a wide variety of selection and/or supply tasks may not only be disconcerting to teachers but may introduce variables that could affect the teacher's score. In addition, the supply-type of item has several disadvantages which reduces its credibility (Wesman, 1971, pp. 90-91):

1. A large preponderance of supply items measure simple facts.
2. It is difficult to phrase supply items so that the same response will be given by examinees.

3. The scoring of supply items is cumbersome, time consuming, and the obtained score is less reliable.
4. Subjective judgment of the scorer diminishes item reliability of supply items.

Another criticism of the available knowledge tests focuses on the variance in the number of items that are alloted to measure knowledge of specific components or aspects of those components. Devoting a disproportionately large number of items to measure a particular skill is an indication of the relative importance of a component as perceived by the researcher. To illustrate, Shannon (1959) used 100 items to test teachers' visual-aural discrimination of phonemes in words. While this may be an important skill, one may question whether 100 items are required to test visual-aural discrimination of phonemes. Similarly, the total test score in Thompson's (1975) instrument is weighted heavily toward knowledge of consonant and vowel sounds. Forty of the 100 items are items on the order of "Mark the word that has the $\bar{a}$ sound that you hear in shady." The allocation of this number of items to decoding did not appear justified considering that, on the average, they were fairly easy and, therefore, were not useful in discriminating among examinees. In her test of reading skills, Broman (1962) devotes 20 of 34 items to knowledge of skill sequences as they are taught in basal reader tests. This is yet another example of the over-reliance upon authors of basal reader texts as authorities of correct practices. In addition, it perpetuates a stereotyped

instructional approach that may not be valid either for groups of learners or for individuals. Thirty of the 85 items in Ramsey's test (1962) require examinees to spell unfamiliar syllables pronounced by the examiner. Symbol-sound relationships of consonants and vowel letters and two-letter combinations such as a, b, d, ch, ng are tested. Results indicated that these items were very easy so based on this evidence their inclusion in a test may not be justified. Further, spelling ability rather than the application of phonic generalizations to help pronounce words is tested.

Whereas some investigators constructed an excessive number of items to measure a component, many authors of tests alloted only one item for this purpose. For example, Thompson (1975) assesses knowledge of a large number of reading skills but the usual procedure is to assess a particular skill with only one item. Similarly, in Durkin's test (1964) most of the ten sections contain only one item to assess a particular phonics skill. Providing only one item to assess a particular skill does not seem sufficient to determine whether or not a person possesses knowledge of the skill, and consequently could limit the validity of the instrument.

Because Artley and Hardin (1975) fail to designate the items that were constructed to measure the seven content areas they identify, the contribution of specific items to a knowledge competency in each of the areas cannot be determined. Therefore, the diagnostic value of the instrument is restricted. In addition, the content validity of the test is limited by the inadequacy of the

universe specifications. Cronbach (1971) states that the requirement for content validity is that the universe boundaries must be well defined so particular items can be included or ruled out by definition (pp. 455-456). In the case of Artley and Hardin's test, one cannot ascertain if the items adequately sample the content categories. If the seven content areas had been specifically and precisely defined, qualified persons would agree whether or not particular items should be included in the test. Consequently, the relevance of the particular items that require examinees to identify an author of a book for children, a theme of a children's story, and the name of an annual book award cannot be established. In the same vein, Mallet's (1972) assumption that the content emphasized in his 20-item test is relevant for first grade teachers and thus is a valid indicator of their knowledge of reading instruction does not seem justified in this author's opinion.

Finally, a number of other factors detract from the quality of the available instruments. First, a large proportion of the test items emphasize rote responses. Examinees are usually required to define a term or state a rule with no attempt to relate it to instructional procedures. Second, when teachers are asked to indicate syllable division one can only hope that syllabication is taught properly and involves more than drawing lines between syllables. Third, the measures of diagnostic proficiency focus on a rather narrow range of diagnostic assessments, and consequently, a narrow interpretation of a teacher's problem solving role. For

example, most of Wade's (1960) problem solving tasks deal with a teacher's ability to analyze student's word attack skills. Areas such as reading comprehension, reading expectancy, and the directed reading lesson are not included. In addition, the instruments do not tap knowledge of the basic underpinnings of the various tests or procedures. For example, construction or implementation of cloze procedures, informal reading inventories, or criterion-referenced tests are not assessed nor are the pros and cons of standardized tests examined. It would seem that this knowledge is foundational if one ascribes to the diagnostic approach in the classroom. Fourth, test items may not be tapping what they were intended to tap. For example, irrelevant clues in the stem may give the answers away or the ability to read charts may be tapped instead of meaningful interpretation of test data (e.g., Thomas, 1975). Or, insufficient cues and misleading directions may prevent an examinee from answering correctly (e.g., Wade, 1960).

While the available tests can be criticized on a number of counts, in this author's opinion the major criticism is the overconcern with <u>what</u> to teach and the lack of focus on <u>how</u> to teach reading. The assumption seems to be that if teachers know content they will be effective in the classroom. Although knowledge of what one is teaching is an important component, knowledge of the process of teaching reading is an equally important component. The failure to assess this component would indicate that one has little basis for assuming competency in the act of teaching. In addition,

the lack of agreement among test authors as to what constitutes the knowledge base in the available tests leads one to conclude that teachers may have differing conceptions of content that is most important for learning to read. Based on this, a need exists for an instrument that taps a knowledge base that is directly related to the act of teaching. A knowledge base that defines a subskills approach to teaching reading would seem important for a number of reasons.

Justification for Tapping Knowledge of a Subskills Approach

Justification for tapping knowledge of a subskills approach to teaching reading is presented in the form of five general premises with selected literature to support each premise. The literature is reviewed and discussed to substantiate that:

1. Reading is a complex skill composed of specific subskills.
2. A skill-hierarchy model provides a valid tool for viewing reading as a system of interrelated and sequenced components.
3. A subskills approach provides a framework for individualizing instruction.
4. A subskills approach contributes to general reading performance.

5. Direct explicit instruction, which is inherent in a subskills approach, facilitates learning to read.

1. Reading is a complex skill which is composed of specific subskills.

Although there is a fair degree of consensus that reading is a highly complex skill which depends upon a series of specific subskills there is little consensus about what these skills are (Stennett, Smythe & Hardy, 1975). This may be due, in part, to the lack of agreement among authorities about how reading should be defined and further confounded by the fact that skills important for learning to read are unimportant for mature reading (Smith, et al., p. 51). Nevertheless, during the last three decades reading experts have put together a number of scope and sequence lists of reading skills that are based on either historical, consensual, or limited empirical support (e.g., Wisconsin Design for Reading Skill Development, Wisconsin, 1971; Communication Skills Program, Southwest Regional Laboratory, 1972). These lists vary in the number and sequencing of skills but they all attempt to break the complex act of reading into specific teachable components. Gray (1960) provided a model for these lists when he identified a comprehensive list of skills common to most reading activities and presented them under four headings:

1. Word perception, which includes pronounciation and meaning,
2. Comprehension, which includes literal and implied meanings and significance and purpose of the author's communication,
3. Reaction to and evaluation of what is read
4. Assimilation -- fusion of ideas acquired through reading with previous experience.

While he recognized that these four aspects of reading operate simultaneously and that reading must be considered a unitary act, his analysis identified component skills for which sequential instruction should be provided.

Theoretical support for subskills was offered by Samuels (1974) when he examined the controversy over wholistic or subskill methods for beginning reading instruction. He states:

> While agreeing with the critics of the subskill approach that too much emphasis can be placed on these subordinate skills, the critics probably are in error in failing to recognize the importance of subskills in the developmental sequence of skill attainment. Just because fluent readers are able to access the meaning of a printed page is no reason to believe that beginning readers can do the same or that we can transfer the sophisticated strategies of the fluent reader to the beginning reader. While it is true that sophisticated strategies can be taught to the less sophisticated, these transfers of skills have been accomplished by doing a task analysis of the sophisticated strategies and teaching these subskills to the beginner.
>
> As the advocates of the holistic approach point out, the essential element of reading--deriving meaning--is destroyed by taking a whole and breaking it down. However, . . . current research suggests that before

> one deals with wholes, smaller aspects have to be mastered first. For example, before one can visually process letter clusters as a unit, individual letters have to be unitized. The controversy between letter-by-letter and whole word processing in word recognition seems somewhat resolved now that we have evidence to indicate words can be processed by fluent readers as a unit while unfamiliar words tend to be processed letter-by-letter. (pp. 21-22)

Over the past half century a number of empirical studies have been conducted which provide support for the assumption that reading comprehension is composed of discrete skills. The first important study on the nature of mental skills involved in comprehension was published by Thorndike (1917). Some of the important early attempts to isolate components of comprehension were correlational studies carried out by Richards (1929), Berry (1931), Touton and Berry (1931), Dewey (1935), and Traxler (1941). Lists of skills produced by these studies ranged from Dewey's list of two skills to Berry's list of sixteen skills.

The earliest factor analysis of reading comprehension skills was conducted by Feder (1938). A number of factor analytic studies have been subsequently reported with reviews provided by Suchor (1959), Chester (1971), and Follman, Lowe and Pfost (1972). Probably the best known of all the investigations of this type was conducted by Davis (1944). In 1968, Davis reinvestigated the unique variance of subtests of reading comprehension and provided some of the strongest support for a multi-factor theory of comprehension. He concluded that "comprehension among mature readers is not a

unitary mental skill or operation" (p. 510). His factor analysis indicated five identifiable skills: (1) memory for word meaning, (2) drawing inferences from content, (3) finding answers to questions answered explicitly or in paraphrase, (4) recognizing a writer's purpose, attitude, tone, and mood, and (5) following the structure of a passage. Research conducted by Spearritt (1972) confirms Davis' (1968) findings with the exception of the skill "finding answers to questions answered explicitly or in paraphrase." Chester (1972) concluded that although research results differ regarding the number and specificity of subskills of comprehension, a review of representative studies provides:

> . . . a growing amount of evidence for the multi-skill theory of comprehension. The majority of the research indicates that most of the measurable variance in tests of reading comprehension can be accounted for in terms of a small number of factors, perhaps six to eight. (pp. 261-262)

2. A skill-hierarchy model provides a valid tool for viewing reading as a system of interrelated and sequenced components.

According to Gagne (1960), a particular skill can be learned if each subordinate skill for the skill above it is acquired by the learner. Then, possession of the relevant skills eventuates in the terminal skill. By using the technique of task analysis, components that are contributors to the task are identified and then instruction can be designed optimally so people become competent in performing the task.

Athey (1974) discussed the relevance of hierarchy systems in

reading instruction in a National Institute of Education position statement. She states:

> Every system of teaching reading presupposes some kind of hierarchy, explicitly or implicitly as a condition of proceeding toward some goal in a rational manner. . . . This procedure has received some theoretical support in the concept of developmental stages and the associated concept of readiness.
>
> On the other hand, hierarchical systems appear to have some basis in the behavioristic approach as exemplified in programmed instruction, in which a terminal task is analyzed into small steps of progressive difficulty and the student is moved through the individual steps at his own pace. The underlying assumption is that learning will progress maximally if the task is broken down into small steps through which the child progresses in an orderly sequence at his own pace.

Further support for a hierarchy model was offered by both Gibson (1965) and Dahl and Samuels (1977). Gibson suggested that there is a sequence from simple to complex in the process of reading acquisition. She distinguished three stages which follow each other, although there may be some overlap. The stages of Gibson's sequential model are:

1. Learning to discriminate graphemes by detecting their distinctive features
2. Learning to decode letters to sound
3. Using progressively higher-order units of structure

Gibson's (1970) analysis of the distinctive features of letters of the alphabet and the way in which awareness of these distinctive features develop suggested that some features may have priority in

the discrimination process and perhaps developmentally, which has implications for instructional sequencing. Dahl and Samuels (1977) developed a four-stage hypothesis/test model which represents the processes involved in reading. Using a stepwise technique, beginning readers are trained in seven component skills which comprise the word recognition strategy of mature readers. The procedure which is recommended is to begin training at the oral language level and when mastery is achieved at this level to introduce the same skill visually.

Empirical support for the hierarchy model has been provided in studies conducted by Rystrom (1970), Guthrie (1973), and Filip (1975) applying the hierarchy model to reading subskills.

Rystrom (1970) suggested that comprehension can be defined as: vocabulary, syntax, item recall, item sequence, inference, and evaluation. He constructed a 57-item test to measure these reading comprehension skills and administered it to remedial groups and non-remedial groups of fourth graders. While there were differences between groups, it was one of degree rather than type. Every group had least difficulty with item recall, while identifying the appropriate sequence of items was difficult for every group. Since differences in difficulty among the skills was marked it was concluded that the components are not simply different measures of the same singular skill. The components were also highly correlated with one another, indicating that if a reader was superior in one component he/she was superior in the other components. Because

some of the skills were not identical in strength and because the intercorrelations were high the systems model which views reading as associated components of skills was supported.

Guthrie (1973) constructed criterion-referenced tests of eight phoneme-grapheme association skills and analyzed the intercorrelations for normal and disabled readers. He found that the intercorrelations among subskills were highly significant for normal readers but low for the disabled group. This suggested that skills are distinct components that are highly integrated in normal children, thus making reading seem to be but one skill. Guthrie concluded that interfacilitation among subskills is necessary for normal reading and that one source of disability among poor readers is the lack of integration and interfacilitation among subskills into higher order units. Guthrie stated that "it is probable that there is a hierarchy of subskills from simple to complex that are learned at different times in the normal acquisition of reading" (p. 17).

Filip (1975) tested the hypothesis that a hierarchy from simple to complex exists among six of the eight grapheme-morpheme correspondence measures developed by Guthrie. Filip concluded that the six skills seemed to fit a hierarchical order with recognition skills generally less complex than production skills, with the exception of single letter production, which appears to be the least complex skill. The hierarchy from least complex to most complex is single letter production, initial letter recognition, consonant cluster recognition, nonsense word recognition, consonant cluster

production, nonsense word production. The fit of the model seemed to be somewhat better for slow than for normal readers.

In summary, although an empirically derived or validated hierarchy of reading skills does not exist, empirical support for the hierarchy model has been provided by studies in which it was applied to reading skills. Additional support for this model is offered by Mason (1977) in a review of studies that have implications for ordering skills learning based on the emergence of linguistic and cognitive abilities. Further, techniques of task analysis or some other logical analysis of reading provide the basis for efficient instruction. Finally, every system of reading presupposes there is a sequence from simple to complex in the process of reading acquisition. Smith, et al., (1978) state:

> Practioners [also] know that skills will be taught in some kind of sequence simply because they can't all be taught at the same time. . . [thus] it seems sensible to be satisfied with sequences that are worked out with regard to logical order, available research, and the practical demands of the classroom. (p. 49)

3. A skill-centered approach provides a framework for individualizing instruction.

In teaching reading, individualization of instruction can proceed most efficiently when teachers help learners focus on what they need to know. Otto (1977) provided support for teaching skills when he states:

> . . . We need them [skills] in order to focus instruction. Or to individualize instruction. . . whatever terminology we use to describe a process whereby we (a) decide what we want learners to know, (b) find out what individuals do and do not know, (c) teach individuals--probably with other individuals who have the same need--what they don't already know.

He further stated:

> Many authors who write reading methods texts say little or nothing specific about skills, yet they rhapsodize about individualization and urge their readers to take a diagnostic approach to teaching. Such urging is like telling people to shoot at a target without saying what or where the target is --or whether the shooting should be done with a shotgun or a rifle. The identification of specific skills brings precision to diagnosis and focus to instruction. (p. 195)

Samuels (1974) observed that individualization of instruction is more manageable if teachers "assume on a priori grounds that there are certain subskills that beginning readers require [and that] these should be taught routinely to students" (p. 23). He pointed out that by teaching the prerequisite skills before a problem appears the number of students who experience difficulty is reduced. In addition, the logistical and managerial problems that are the outcome of providing individual help upon the child's request or apparent need are avoided. Then, individualization of instruction is workable in terms of efficient use of teacher time in helping individual students acquire skills before they bog down in the reading process.

Finally, Otto, et al., (1974) proposed that teachers must have a way to organize instruction if they are to do an effective job of helping learners focus on what they need to know. They stated that

individualization of instruction "can proceed more efficiently when we have a way to assess the individual's strengths and weaknesses and to adapt our teaching in view of what we know about each individual." To facilitate the teaching/learning process they developed the following guidelines:

1. Identify essential skills: Decide which skills are essential to successful reading performance.
2. State objectives: Describe each skill in terms of the behaviors that are expected of a student who has mastered it.
3. Examine individuals' skill development: Assess the skill-development status of students with criterion-referenced tests and informal performance tests.
4. Identify and carry out appropriate teaching/learning activities: Use any methods or materials that offer appropriate learning opportunities at a pace that is appropriate for the student.
5. Evaluate the results: Determine if the skill-related objectives are being reached, accompanied by functional reading ability. (pp. 39-41)

4. A skill-centered approach contributes to general reading performance.

Experimental evidence supports reading programs that place an early emphasis on skill development rather than a whole word

approach that features drill and repetition of high frequency words and reliance on context clues to identify words. Reading programs which promote greater reading achievement include basal reader texts of the whole word type that use a phonics strand that may be relatively independent of the text, letter-sound controlled texts, or skill-centered reading programs.

Bond (1966), in analyzing data from the First Grade Reading Studies, indicated that:

> . . . no matter what the underlying method is, word study skills need to be emphasized and taught systematically. This is best shown by the superiority of approaches which augmented the basal reader with a phonetic [phonic] emphasis as compared with basal readers as usually taught. (p. 9)

Similarly, Chall (1967) and Harris (1969) independently analyzed the existing experimental comparisons of meaning emphasis versus code emphasis approaches and pointed to the highly probable superiority of beginning reading programs in which an essentially code emphasis method is used. She found that an early emphasis on code learning not only produced better word recognition and spelling but also made it easier for the child to eventually read with understanding. Harris (1969) stated that:

> To place the prime instructional emphasis upon the direct translation of words into their meanings without attention to the intervening variables of letters represents a fundamental misconception of the means and ends of initial instruction in learning to read as alphabetic language. (p. 1980)

Stallard (1977) noted that evaluations of skill-centered reading programs have been conducted to investigate student gains in

overall reading ability, as demonstrated by students' scores on standardized reading achievement tests. She stated:

> The following reports cite evidence of gains in overall reading ability resulting from one or more of these programs, as measured by scores on standardized reading tests: Aaron (1974), Horn (1974), Rosenkranz (1973), Klopp (1976), Hubbard and Sals (1975), Green and Roudabush (1975), and Cohn (1974).

Further support for skill-centered reading programs was provided by Askov, Otto and Fischbach (1971) and Blohm (1976, 1978), who indicated that mastery of reading skills is positively correlated with level of reading achievement.

Askov, et al., (1971) examined the relationship of word attack mastery to the general reading achievement level of students in grades one through six. A statistically significant relationship between word attack mastery level and reading performance as measured by general tests was found at all grade levels.

In his first study, Blohm (1976) explored the relationship of skill mastery of fourth grade subjects to general reading achievement. He noted the subjects' word attack mastery levels along with their scores on the Metropolitan Reading Test. Results indicated that skill mastery is positively correlated with level of reading achievement.

In a second study Blohm (1978) examined the effect on general reading comprehension of teaching and testing four main idea subskills with 157 tenth grade English students in three treatment groups. The instruction group was taught and tested for mastery of

the main idea subskills while the testing group received only mastery tests and the control group received no specific treatment during the eight week period. General reading comprehension was assessed immediately at the end of the eight week period and one month later. A combination of teaching and testing main idea subskills resulted in a sustained positive effect on general comprehension.

In view of the evidence, reading skill instruction that is systematic provides the means for moving children to independence in their reading comprehension. Then, if the attainment of basic reading skills is accompanied by the ability to cope with reading tasks the ultimate goal of reading instruction is achieved.

5. Direct explicit instruction, which is inherent in a subskills approach, facilitates learning to read.

Otto, et al. (1974), Rosenshine (1977), Berliner and Rosenshine (1977), and Duffy and Sherman (1977), emphasize the importance of providing direct instruction in basic reading skills.

Berliner and Rosenshine (1977) state that an emerging body of literature "indicates that teaching behavior focused on direct instruction results in increased acquisition of student knowledge and skills" (p. 382). In direct instruction of basic reading skills "the teacher controls instructional goals, chooses materials appropriate for the student's level, and paces the instructional episode" (Rosenshine, 1977, p. 9a). Students are brought into

contact with the content of the curriculum and kept in contact until the knowledge is acquired (Berliner and Rosenshine, 1977).

Berliner and Rosenshine (1971) cite and describe a number of research studies which support components of direct instruction. Instructional variables that are directly related to making progress in reading include seven essential aspects (pp. 375-396):

1. Instructional goals for both instructional episodes and seatwork are clearly stated.
2. Instructional time is sufficient as well as continuously focused on the lesson within a given time period.
3. Students work in groups under adult supervision rather than in a group work or independent activities that are monitored infrequently.
4. The amount and range of content covered is maximized.
5. Questions at lower cognitive levels that produce many correct answers (e.g., academically focused questions, questions with single answers, drill) are used with primary age children.
6. Immediate and academically focused feedback is provided.
7. Teaching behavior is directly aimed at furthering academic achievement in an environment that is characterized as structured, orderly, democratic,

and warm.

Otto, et al. (1974) state that the main task of the reading teacher is to help each child master the skills as rapidly and efficiently as possible. They propose that efficient learning is encouraged when teachers help children focus on what they need to know. The first step in providing focused instruction is to identify the skills that are essential to success in reading and to state these skills in behavioral terms. Then, based on a description of behaviors considered relevant to each skill the teacher has a basis for determining pupils' skill development status, organizing instruction, and evaluating the results. When the teacher makes the facilitation of learning the main function of teaching, reading skill instruction that is systematic will culminate in their mastery.

Finally, Duffy and Sherman (1977, p. 43-55) describe an instructional model for teaching reading skills. They describe six specific ways to directly assist the learner that should be included in every instructional episode:

1. Creating a response set: The teacher must acknowledge the difficulty of the task and that learning is seldom a one-time affair.
2. Directing attention: The teacher must state what is to be learned and must emphasize the task's salient features.
3. Highlighting: Salient features must be highlighted so that the learner focuses on the critical features

of the task.

4. Diminishing the assistance: Highlighting assistance is diminished as the learner begins to respond correctly.
5. Response: Frequent responses are elicited during a lesson.
6. Reinforcement: Reinforcement should immediately follow the correct response.

Summary

An examination of the available instruments indicates very little agreement as to the knowledge base that is essential for teaching reading. Because the tests are unfocused, knowledge of the content of reading instruction has not been systematically examined and the extent of teachers' knowledge about the teaching of reading cannot be determined. Consequently, whether or not a teacher is "knowledgeable" is a decision that rests on the subjective judgment of the test author with respect to the particular components he/she chose to focus on, the process by which the composition of the test base was determined, and the number and type of items that were constructed. The implication inherent in the instruments is that if teachers know a particular set of terms, generalizations, diagnostic procedures, etc., they have the basis for effective instruction. Coupled with the lack of consensus concerning what content of reading instruction is most important is the absence of

instruments for assessing knowledge of how to teach reading. Thus, it would seem important to attempt to measure a knowledge base that is directly related to competency in the act of teaching reading. A knowledgeable teacher would then be defined as one who has the requisite knowledge to facilitate children's learning to read, based on knowledge of what children need to learn and on how to implement efficient instruction.

Since the systematic teaching of specific skills, which culminates in the acquisition of these skills, is an integral part of the process of reading instruction, knowledge of a subskills approach would appear to provide a foundation for effective reading instruction. An explicit focus on skill development not only compliments a language experience, basal reader, or individualized approach but also provides the means for moving children to independence in their reading comprehension. Otto (1977) pointed out that combining skill instruction with useful elements of basal reader and language experience approaches allows teachers to handle the instructional needs of students. In the same vein, Athey (1977) commented that "even the most ardent advocate of a language experience approach will find it necessary to do some skill-building in the classroom so why not make it systematic?" (p. 233).

Further, a subskills approach to teaching reading is viewed as pragmatically valid because it provides a framework for individualizing instruction while allowing teachers to exercise their professional expertise and initiative (Otto, 1977; Athey, 1977; Smith,

Otto, and Hansen, 1978). Based on the identification of essential skills, teachers differentiate and pace instruction for individual students with appropriate materials and procedures.

Considerable support for a subskills approach to teaching reading establishes its relevance for making the main function of teaching reading the "facilitation of learning." Specific justification for the instrument has been provided with theoretical, empirical, and consensual support which points to the efficacy of a subskills approach to teaching reading. Evidence has been cited which supports the assumption that a complex skill such as reading can be acquired if smaller aspects, or subskills, are mastered first. In addition, a hierarchy system of relevant skills which is "worked out with regard to logical order, available research, and the practical demands of the classroom" (Smith, et al., 1978, p. 49) is basic to proceeding toward the goal in a rational manner. Further, guidelines have been presented (Otto, et al., 1974) which provide teachers with a framework for optimally designing instruction so that learners can focus on what they need to know and can become competent in general reading performance. Finally, in view of this evidence it would seem that if a teacher possesses knowledge of a subskills approach she/he will possess a core of knowledge that is pedagogically valid in that efficient instruction can then be implemented. Based on this rationale, it appears that the construction of an instrument to measure knowledge of a subskills approach to teaching reading is justified.

Statement of the Problem

To date there exists no means to assess teachers' knowledge of a subskills approach to teaching reading. With such an instrument it would be possible to predict pre-service teachers' success in reading coursework and might also prove valuable as a diagnostic instrument for the development of in-service training of teachers in a subskills approach. In addition, the instrument could be utilized as a research tool to assess the correlation between knowledge and a variety of factors.

Based on the indication of the need for the instrument, an instrument was constructed to assess elementary teachers' knowledge of a subskills approach to teaching reading. The particular sub-problems which directed the project were:

1. Developing an instrument that possessed content validity.
2. Developing an instrument that was sufficiently reliable for the instrument's intended uses.

In addition, a comparison of the differences between the group mean scores for teachers with zero or one reading course and teachers with three or more reading courses was made. Then, the correlations between number of reading courses and years of teaching experience to test scores were examined.

Chapter II reports instrument development procedures which includes developing the test specification, item selection, field

tryouts, data analysis, and instrument revision.

CHAPTER II

INSTRUMENT DEVELOPMENT

The construction of an instrument to measure knowledge of a subskills approach to teaching reading involved these procedures:

1. Developing the test specifications
2. Constructing and selecting items for the preliminary test form
3. Field Tryout I, analysis, and revision of the preliminary test form
4. Administering and analyzing the revised test form.

Decisions that were made which relate to a particular phase of test development and specific evidence which supports these decisions are presented.

Developing the Test Specifications

The planning stage involved making a number of decisions to ensure that subsequent phases of test construction would proceed on a sound basis. These included: (1) stating the purpose of the test, (2) developing the test outline, (3) selecting the item format, (4) determining the number of items to be included in the pilot test, and (5) establishing the criteria for item evaluation.

Purpose of the Test

The purpose of the test is to measure the extent of elementary teachers' knowledge of a subskills' approach to teaching reading. Elementary teachers are defined as teachers in grades one through six. The test scores resulting from administration of the instrument will be used for evaluating total achievement of individual teachers.

Test Outline

The outline of test content evolved from the review of literature in Chapter II in which justification for tapping knowledge of a subskills approach to teaching reading was provided. First, a definition of a subskills approach emerged which provided a conceptual framework for the proposed instrument. The definition that emerged and which appeared to present professional consensus (Otto, et al., 1974; Samuels, 1974; Athey, 1975; Duffy and Sherman, 1978) was: A subskills approach to teaching reading is reading instruction that is focused on the systematic, sequential development of basic reading skills. The primary aim is to help students make continual progress in mastering needed skills and in attaining functional reading ability (Otto, Chester, McNeil and Myers, 1974). A subskills approach is based on three assumptions (Athey, 1975).

1. Essential reading skills do exist and the subskills can be taught. (Mason, 1977, provides the following qualifiers for this assumption:

-- Skills must constitute observable and measurable behavior which can be viewed within the context of some terminal objectives.

-- Skills are a necessary component of reading.)

2. Progress through an ordered set of skills generally assumes the notion of instructional hierarchies.

   a. Essential subskills may be within larger skills.

   b. Mastery of some subskills is dependent on prior mastery of other skills.

3. A skill-oriented instructional program is pedagogically efficient and effective.

Subsequent to deriving the definition of a subskills approach, the several major categories of content to be covered by the test were identified and defined. In specifying the major content categories in the test outline the following guidelines were observed. First, the categories must focus on a knowledge competency that is directly related to teaching reading as a skill-development process. Second, the categories must indicate the general areas of achievement that are important. Third, each category must be made as maximally different from each other as is possible so a given item tests only one content area.

The following outline of test content was developed upon which the instrument drew.

CATEGORY I: Identifying Skills to be Taught/Learned

A. Teachers identify essential skills. The first step toward organizing reading instruction as a skill-development process is to decide what skills to teach (Otto, et al., 1974, p. 39). The skills that are identified are those that are assumed to be essential to reading competence (Otto, et al., 1974, p. 39). As no validated list of essential skills is available, there is considerable diversity among skill hierarchies (Athey, 1974, p. 17; Duffy, 1978, p. 54). The particular skills that are chosen for instruction reflect logical and limited empirical demonstrations of being essential for successful reading. (Athey, 1975, p. 17).

b. Teachers operationalize skills. To ensure that skills are taught and learned teachers further explicate the skills by specifying what they (the skills) are like (Otto, et al., 1974, p. 40). When specific skills are stated as objectives, descriptions may range from very definitive statements expressed in behavioral terms to those that describe relevant activities for tasks (Eisner, 1969, p. 20). In some instances expected learning outcomes may be predictable while in other instances it may be unrealistic to prescribe outcomes in advance (Eisner, 1967, p. 279;

Otto, et al., 1974, p. 109). When skills are operationalized teachers not only have a vehicle for teaching and assessing them but also for determining the order in which they are to be taught (Otto, et al., 1974, p. 105). Since everything cannot be taught at the same time, logical sequences of learning activities are worked out which reflect thoughtful decisions in terms of their ordering (Otto, et al., 1974, p. 105; Duffy, 1978, p. 521).

CATEGORY II: Promoting Skill Development

A. Teachers systematically teach specific skills.

The skills that are assumed to be those that readers require for successful reading are taught routinely to students with similar needs (Samuels, 1974, p. 23). Students are systematically grouped and regrouped on a short-term basis as needs change as a result of instruction/learning (Otto, et al., 1974, p. 243; Duffy & Sherman, 1977, p. 64). Because skills are not tied to particular grade levels, individual skill development can be both continuous and at a rate that is right for the individual student (Otto, et al., 1974, p. 22, 40). This strategy requires the organization of instructional materials and a continuous record of skills individual students have learned (Otto, et al., 1974, pp. 116-119;

Duffy & Sherman, 1977, p. 69).

B. Teachers provide focused instruction to encourage efficient learning. Skills are not left to develop as a result of generalized reading lessons (Duffy & Sherman, 1978, pp. 6-7). Rather, specific skills are developed in lessons that focus on the skill to be learned (Duffy & Sherman, 1978, pp. 43-55). Efficient learning is encouraged when teachers use appropriate instructional procedures and materials (Gagne, 1974, pp. 97-122).

C. Teachers facilitate remembering. Practice which consists of repeatedly performing the task is necessary if students are to retain what has been taught/learned (Klausmeier & Goodwin, 1975, p. 264; Duffy & Sherman, 1978, pp. 53-54). Practice activities are completed both as and after a skill is taught, with feedback provided by the teacher (Gagne, 1974, p. 115; Klausmeier & Goodwin, 1975, pp. 350-351). Spaced practice, rather than massed practice, results in greater retention (Klausmeier & Goodwin, 1975, pp. 348-351). With some tasks overlearning or drill facilitates remembering (Gray, 1960, p. 9; Samuels, 1974, p. 20; Otto, et al., 1974, pp. 44-45.

D. Teachers arrange activities for learners who differ from each other in terms of proficiency. If the skill has been taught but has not been learned another lesson should be planned and taught with alternative approaches and materials because instruction has been unsuccessful (Samuels, 1974, p. 23). If the skill has been learned the teacher must help the student transfer the learned skill from the instructional situation to reading books and increasing functional reading ability (Otto, et al., 1974, p. 41; Samuels, 1974, p. 22; Duffy & Sherman, 1977, pp. 57, 84). Since basic skills span content areas, providing for lateral transfer leads to functional reading ability (Gagne, 1974, p. 116; Otto, et al., 1974, p. 22; Gray, 1960, p. 22).

CATEGORY III: Determining Skill Development Status

A. Teachers assess pupil skill development before and after instruction. Testing occurs prior to teaching to determine if the student already knows the skill, and following instruction to determine if the skill has been learned (Otto, et al., 1974, p. 114; Duffy & Sherman, 1977, p. 44). In diagnosing a learner's competency in a task, it is better to test his ability on the cumulative task rather

than its prerequisites (Duffy & Sherman, 1977, p. 44). Test exercises must be appropriate for assessing the specified skill (Duffy & Sherman, 1977, p. 44).

B. Teachers use criterion-referenced measures. The measures that are most useful to assess specific skill development are criterion-referenced measures (Otto, et al., 1974, p. 114). The student's achievement is assessed relative to some predefined standard rather than compared to the performance of others (Otto, et al., 1974, pp. 113-114; Duffy, 1978, p. 522). The performance standard may be either an explicitly stated criterion level in a criterion-referenced test or implicit in the form of teacher internalized criteria (Klausmeier & Goodman, 1975, p. 106). Criterion levels for judging the adequacy of performance may be invariant in some instances or may be flexibly applied in others (Klausmeier & Goodwin, 1975, pp. 103-108). For example, a set criterion level seems realistic for word attack skills because of the level of proficiency that is required to handle basic tasks while it seems less applicable to creative reading skills (Otto, et al., 1974, p. 109). Rather, teachers can set that level for individuals contingent on relevant instructional and learner variables (Otto, et al., 1974, pp. 207-208;

> Klausmeier & Goodwin, 1975, p. 106). Finally, criterion-referenced measures may take several forms, depending on the nature of the particular skills and the situation. They may be formal or informal, and they may be written or oral (Otto, et al., 1974, p. 114; Klausmeier and Goodwin, 1975, p. 452).

When the test outline had been conceptualized up to this point, decisions were made concerned with: (1) the relative emphasis of each content category, and (2) the level of specificity of the items. In the absence of any good reasons to the contrary, it seemed reasonable that the total number of test items should be distributed uniformly between the categories. Although assigning equal numbers of items to the categories was considered to be a rough estimate, it seemed appropriate at the present stage of test development. But the intention of the author was not to force a uniform distribution if it seemed inappropriate. With regard to specificity, a mid-level of specificity for the items was established. This meant that items were to be written that would not be too global so as to be vague or ambiguous, nor too specific (e.g., symbol-sound correspondences). Although one might argue that a test that purports to cover the field adequately would include items of varying levels of specificity, it was reasoned that constructing items at a mid-level of specificity would keep the instrument a reasonable length and would allow the author to focus on information that was neither too esoteric nor too

general or unfocused.

After the decisions concerned with relative emphasis of the categories and specificity level of items were made, this information was submitted along with the test outline for evaluation to four persons who were qualified to judge the adequacy of the test outline. The panel of subject matter experts consisted of:

Gerald Duffy, Professor, Michigan State University, East Lansing, Michigan;

Dale D. Johnson, Professor, University of Wisconsin-Madison, Madison, Wisconsin;

Wayne Otto, Professor, University of Wisconsin-Madison, Madison, Wisconsin;

S. Jay Samuels, Professor, University of Minnesota, Minneapolis, Minnesota.

Each evaluator was also given an abbreviated copy of the research proposal and an evaluation sheet to aid in the review. The following questions were asked of evaluators concerning the categories that comprised the test base:

1. Do you agree or disagree that the three categories are appropriate on the basis of their: (a) definitions, and (b) equal emphasis (equal number of items)?
2. Are there other topics that have not been included?
3. Are there topics that should be deleted or revised?

A copy of the evaluation form is in Appendix A.

Responses of the four evaluators indicated that, in their opinion, the selected categories and the relative emphasis of each were appropriate. Additional topics, revisions, or deletions were not suggested. But, the test outline was not viewed as "frozen" because "the experience gained in writing test items often contributes to further clarification of the test objectives, so frequently it becomes desirable to alter the test outline as item writing proceeds." Thus, "the test constructor should not hesitate to make adjustments in the blueprint based on logic and reason, while ever mindful of possible confounding effects" (Tinkelman, 1971, p. 55).

Once the subject matter of the test was established, the form of knowledge that would be assessed by the instrument was specified. Since the form of knowledge should be relative to the purpose of the test, the instrument would focus primarily on knowledge _that_ and to a lesser degree on knowledge _how_. This decision was based on Otto and Harper's (1977) discussion of the relevance of Soltis' (1968) definitions of forms of knowledge for purposes of test construction. Knowledge _that_ refers to cognitive, verbal or propositional knowledge; one knows facts or has information. Knowledge _how_ refers to abilities that are teachable and learnable, where one possesses certain skills and is able to perform certain operations with more or less success. Knowledge _to_ refers to the propensity or disposition to act in certain ways. Otto and Harper conclude that while knowledge _to_ does not lend itself to objective measurement, paper-and-pencil items can probably be constructed to assess the how's of reading pedagogy and the _that's_ which form the basis of knowledge _how_.

Item Format

The decision to assess knowledge that and knowledge how with an objective item format was based on the following reasons (Mehrens & Lehmann, 1973):

1. A paper-and-pencil achievement test, rather than an observational tool, provides a more valid and reliable measure of conceptual knowledge.
2. A paper-and-pencil test has the advantage of being easy to score, and summarize, and it is objective.

Specifically, the multiple choice form was chosen for the test items because it has the following advantages (Thorndike & Hagen, 1969, p. 71; Mehrens & Lehman, 1973, pp. 280-281):

1. Factual recall as well as complex abilities can be measured.
2. Broad sampling of content is possible.
3. Score variations due to guessing are relatively small.
4. Freedom from response sets is maximized.
5. Efficiency in terms of the number of questions that can be asked in a prescribed time and space needed to present the answers is significant.
6. Scoring by persons or machines is rapid and accurate.
7. Providing diagnostic information is a potential outcome.

The decision to write four-choice items was based on the fact that

as the number of answer choices increases from two to four or five the chance level for getting an item correct is reduced (Harris, 1968, p. 5).

Number of Items

Once decisions were made concerning the purpose of the test, what it is measuring, and the appropriate item format, two additional details were considered before item writing began. These were: (1) the number of items to be included in the pilot test, and (2) the establishment of a criterion for item evaluation. The sources that were consulted in an attempt to determine the appropriate number of items revealed that the number of surplus items specified as a margin for discard varies. While Conrad (1951, p. 261) states that the margin of discard adheres to no universally applicable rule, Harris (1968, p. 6) specifies twice the number to be included in the final test form. On the other hand, Tinkelman (1971, p. 74) specifies 10 to 20 percent when items are factual and fewer categories are defined. He also states that the mortality rate is higher for inexperienced item writers. The decision was to write 100 items with 50 to 60 items as the outcome from field testing. The anticipated mortality rate was based on a judgment of the author's level of experience as an item writer as well as the number of categories that were specified and that the items are factual.

Item Evaluation

In establishing the criteria for item evaluation, the strengths and weaknesses of the items were judged based on quantitative evidence regarding the difficulty and discriminating ability of each item in relation to the importance of the item for sampling the content. Because the discriminating power of an item tends to be altered depending on whether a particular group is more or less able or upon the size of the group, this was also taken into consideration in item evaluation. Thus, since it is unrealistic to regard item analysis data as stable properties because they are specific to the particular group on which they are obtained (Furst, 1958, p. 317), an exact level for determination of "good" items was not established. Rather the statistical information for each item was viewed relative to selecting the items on a best basis while still accomplishing the purpose of the test in addition to giving a comprehensive sampling of the content.

In establishing the criteria for the items, statistical output from the GITAP (General Item Analysis Program, Baker, 1969) was utilized. The indices of difficulty level and biserial correlation were used as a pair to judge each item. In determining the appropriate difficulty level for an item, Ebel (1972, pp. 395-401) states that items in a midrange of difficulty discriminate almost as effectively as items of 50 percent difficulty. Using the guidelines established by Ebel, a spread of about .25-.30 percent on either side of the ideal mean was tolerated in selecting items for the

final test form. Items that fell outside this range were eliminated because an item that is so easy that almost everyone can answer it correctly or so difficult that only a few examinees can answer it correctly does not discriminate. In reference to the biserial correlation coefficient, a value of .30 or more for the correct response was desirable because it is at this point that an item discriminates most effectively (Harris, 1968, p. 11). However, items with lower biserial correlation coefficients were retained in the revised test form. Their retention was based on several factors. First, their elimination would narrow the scope of the instrument. Second, (Ebel, 1972, p. 400) the difficulty of writing items that show high discrimination in the field of reading and the fact that a clearly defensible answer is less likely precludes items with lower discrimination indices.

Constructing and Selecting Items

The second phase of instrument development involved constructing and selecting items for the preliminary test form. This included: (1) identifying topics and content for the items, (2) writing items, and (3) reviewing and revising the item pool.

Identifying Topics and Content

Because the test outline did not specify specific topics or content for individual items, it was necessary to first: (1) derive a list of topics from the test outline on which items might

be based, and (2) identify content that would be translated into items. Since the purpose of the instrument was to measure a knowledge competency that was related to teaching reading as a skill development process, the topics focused on knowledge that seemed foundational to this approach. The specific topics which emerged and which were subsequently utilized in the preparation of test items were:

CATEGORY I

Teachers know that the first step in organizing instruction is to decide what skills are important for success in reading.

Teachers know essential skills.

Teachers know about behavioral objectives.

CATEGORY II

Teachers know that skills are taught routinely, at the student's rate of development.

Teachers know that skill instruction is usually provided to short-term groups with similar needs.

Teachers know why records of skill development are kept.

Teachers know about appropriate instruction of specific skills.

Teachers know about practice.

Teachers know about transfer from skill lessons to reading texts.

CATEGORY III

Teachers know how to assess skill development.

Teachers know about criterion-referenced testing.

In identifying content that would be translated into specific items, a given set of textbooks, reading methods texts, and related materials was studied (see Appendix B). The criterion for selecting the particular set of source materials was that they were representative of those teachers encounter in their pre-service and in-service coursework. The relevance of content for building specific items was based on whether it: (1) measured the three categories, and (2) represented a fair degree of consensus among authorities.

Writing the Items

After the item content was identified, one hundred items were developed which exhibited the following characteristics:

1. Items were the four-choice, multiple choice type.
2. Item stems were stated as incomplete statements or as questions.
3. The negative stem was used sparingly.
4. "None-of-these" of "all of the above" options were avoided.
5. All items were independent (e.g., a best answer to one item did not depend on the answer to a previous item).
6. The "best-answer" instead of "correct-answer" responses were developed.
7. The items were as appropriate in technical considerations as the author could make them (e.g., clear

unambiguous language, no clues to the correct answer, no editorial incongruities).

The alternatives "none of these" and "all of these" were avoided because of the inherent ambiguity in these devices. Both of the alternatives are inappropriate in "best answer" items. For example, an examinee cannot reasonably be asked to mark "none of these" when the general directions state that the best of several imperfect options is to be selected (Wesman, 1971, p. 117; Ebel, 1972, p. 199).

In reference to the answer type, the "best answer" response was chosen because the complex nature of the subject area predicates the possibility that competent experts might not agree on the correct answer. In discussing this point, Cronbach (1971) states:

> One cannot hope to perfect items to the point where a hypercritical reviewer cannot quibble over conceivable ambiguity or exceptions to the keyed answers. Subjects dealing with complex subject matter are invariably open to some misinterpretations as any textbook writer can testify. (p. 457)

Although the items appeared to be adequate in terms of measuring the content specified in the test outline, comparatively fewer items were written for Category III. Thirty-six items were written for Category I, 39 for Category II, and 25 for Category III. The specification of equally distributing the items among the categories was not met for several reasons. First, it had become apparent that a larger number of items were required to assess Category II. This observation was based on two factors: (1) six

topics relating to teacher knowledge were identified for Category II which was at least twice the number for each of the other two categories, and (2) the content emphasized in this category seemed most essential for teaching reading as a skill development process. The particular content was concerned with management of instruction as well as instruction, practice, and transfer of specific skills.

Reviewing and Revising the Items

Prior to the first field tryout, the items were subjected to two reviews to determine their adequacy for measuring the content specified in the test outline. Subsequent to each review items were revised, deleted, or new items were written based upon the reviewers' suggestions.

Initial Review. Once written, the items and the test directions were subjected to a critical review by four persons, who because of their educational background and professional role, were qualified to judge the adequacy of the items in terms of subject matter content and technical aspects. Each of the reviewers possessed a master's degree or more in reading and had elementary teaching experience. Two of the reviewers had developed objective-based teaching materials and tests to measure reading objectives and one reviewer had worked as a reading specialist in charge of implementation of the Wisconsin Design for Reading Skill Development. Each reviewer was furnished with: (1) a copy of the items arranged in categories that corresponded to those in the test outline, (2) a

copy of the test directions, (3) explicit instructions for reviewing the items and directions, and (4) a copy of the test outline (see Appendix C). Review procedures primarily followed the guidelines set down by Tinkelman (1971). Reviewers determined how appropriately the items sampled the content categories defined in the test outline and indicated whether there was agreement as to the best answer for each item. The technical aspects were concerned with item form and adherence to measurement principles. Item stems were examined to determine if they were free from ambiguity, communicated clearly, and did not provide clues to the answer. With regard to the directions, reviewers determined if they: (1) contained technical defects, (2) were clear, and (3) were explicitly stated.

Subsequent to the review of the items, an analysis of the consistency of the judges' responses to the following questions was performed with Cronbach's Alpha (Cronbach, 1951, pp. 297-334):

1. Is the item keyed to the appropriate category?
2. Do you agree that the best answer is the best answer?

The judges tended to agree to a high degree because Alpha for the first question was .85 and .91 for the second question.

The review was supplemented with a session in which the reviewers discussed with the author objections to particular items and offered comments that were used in the preliminary screening and revision of the item pool.

There appeared to be a general feeling among the reviewers

that the items that tested knowledge of comprehension skills needed to be replaced. The two main sources of difficulty were: (1) several options were best answers, and (2) basic definitions of comprehension terms and clear differentiation between them are lacking in the literature. Other comments identified items that exhibited the following characteristics:

1. Very specific information was tested.
2. Differences of opinion among subskills authorities concerning issues was tested.
3. Jargon or terms whose meaning might be unknown to the examinees was the source of difficulty in an item.
4. Lack of qualifiers in the item resulted in an unclear or unfocused item.

Then, based on the comments and suggestions of the reviewers, deficiencies in the items and directions were eliminated as nearly as possible. In the process, new items were written and items were revised and deleted. The end result was a total of 101 items with 33 items in Category I, 42 items in Category II, and 26 items in Category III. Although the categories were unbalanced in terms of number of items, the author decided that rather than eliminate items at this point, data furnished later in the test development process would be the determiner.

Second review. A further assessment of content validity was made when the best answer for each item was independently selected

by 12 persons who formed a heterogenous group in terms of number of reading courses and years of teaching experience. A full statistical analysis was not performed but, rather, the summary statistics and distribution of responses for each item was computed and studied. The scores ranged from 62 to 89 items correct with a mean score of 74 and a median score of 73.5. Five of the six people who had the most reading courses (10 or more) scored above the mean.

Seven of the persons were individually interviewed to furnish information for improving the items and directions. Their comments helped identify items that: (1) had two answers that seemed to be best answers, (2) were difficult in terms of vocabulary, or (3) did not communicate clearly or fully. Based on their comments, items were revised or deleted, and similar items were combined. The end result was a total of 86 items. Twenty-eight items measured Category I, 42 items measured Category II, and 16 items measured Category III (see Appendix D). While the categories could have been equalized by either padding Categories I and III with items or cutting down the number of items in Category II, both these options seemed unwise. With padding, knowledge that was less relevant would be tested, while eliminating items would result in an inadequate measure of Category II. In addition, the basis for designating unequal numbers of items to categories had been provided when the original weights had been cited as rough estimates that might be adjusted as the test was being built.

As a final step in readying the instrument for the first field

tryout, the items and directions were edited to correct inconsistencies in overall format and awkward wording. The editing was done by a person who was knowledgeable in the subject matter area and had worked professionally as an editor. The items were then rewritten to incorporate the suggestions of the editor and were assembled, along with the directions, into the preliminary test form (see Appendix E).

Field Tryout I, Analysis, and Revision of the Preliminary Test Form

The third phase of instrument development involved: (1) securing experienced teachers as examinees, (2) administering the instrument in a field tryout, (3) analyzing the resulting data, (4) revising the preliminary test form, and (5) reassessing content validity.

Examinees

Because the instrument was intended for teachers of grades one through six, the sample for the first field tryout was representative of the intended population. The examinees consisted of fifty-seven experienced teachers who were selected on the basis of availability and willingness to participate.

Forty-five of the examinees were teachers in a Westcentral Wisconsin school district that had a pupil enrollment of 10,460 for the 1978-79 school year and twelve examinees were experienced

teachers enrolled in reading coursework at the University of Wisconsin-Madison. Demographic data related to training and experience were collected for the purpose of piloting the personal data form and describing the examinees in the trial administration sample. Data for the trial administration sample indicated that training in reading courses ranged from 1 to 10 with the mean number of courses 3.63, the median 3, and the mode 2 courses. Years of teaching experience ranged from 1 to 37 years. Although 41 of the teachers had taught 10 years or less, the mean years of teaching experience was 9.56 years. Twenty-two teachers taught at the primary level while 35 teachers taught at the intermediate level.

Administration

The instrument was administered to small groups of teachers by the author or by two other examiners. No time limits were imposed. The test was usually completed in 45 minutes although some examinees completed it in 30 minutes and one examinee required almost two hours. To minimize the score variance that results from factors external to the examinees, an attempt was made to administer the test under uniform conditions. First, written instructions for administering the test were followed in each of the testing sessions (see Appendix F). Second, rooms were scheduled for the testing sessions that were free from distracting noises. Third, teachers were not interrupted during the testing period.

Analysis of the Data

The GITAP program was utilized to analyze the data obtained from the first field tryout. The program supplied: (1) summary statistics for the total sample, (2) the reliability coefficient of the instrument, and (3) the difficulty level and biserial correlation coefficient of each item (see Table 2).

Scores of the test group ranged from 39 to 80 items correct, with a mean score of 55.8, a median of 55, and a standard deviation of 7.64. The standard error of measurement was 3.83 and the Hoyt internal consistency reliability coefficient was .75. An examination of these summary statistics revealed that the test was rather efficient, which means that the items are of appropriate difficulty. The average score of 55.8 is near the ideal mean score of 53.6, which is the midpoint of the range between the highest possible and expected chance score. Because the mean score of a test is determined by the mean difficulty of the items composing it, a mean score that is much higher or lower than the ideal mean would indicate that the test is inefficient. Thus, one can conclude that the instrument does not waste the examinee's time trying to respond to items that almost no one can answer correctly or reading and responding to items that almost everyone knows (Ebel, 1972, p. 376). In addition, the standard error of measurement is small, which indicates that two-thirds of the time a teacher's true score lies within a range of 3.83 points above and below the obtained score on the instrument. Finally, the Hoyt internal consistency reliability coefficient was

Table 2

Difficulty Levels on a Biserial Correlation

Coefficient Values for Items in Field Tryout I

Item	Difficulty Level	Biserial Correlation Coefficient
1*	.39	.46
2*	.56	.20
3*	.89	.25
4	.96	.46
5*	.70	.49
6*	.47	.46
7	.95	.01
8*	.61	.53
9	.16	.68
10*	.65	.37
11*	.61	.35
12	.93	.21
13	.82	-.05
14	.30	.08
15	.26	-.05
16*	.67	.31
17	.93	.21
18*	.53	.17
19*	.51	.26
20*	.81	.26
21*	.81	.37
22	.92	.07
23	.26	.14
24*	.53	.24
25*	.77	.35

Table 2 (Continued)

Item	Difficulty Level	Biserial Correlation Coefficient
26*	.75	.13
27*	.67	.42
28*	.77	.37
29*	.79	.40
30*	.63	.22
31*	.32	.47
32*	.68	.53
33*	.47	.47
34*	.70	.20
35*	.77	.28
36*	.33	.22
37*	.91	.52
38*	.88	.30
39*	.60	.52
40*	.65	.55
41*	.82	.25
42	.25	.30
43*	.67	.27
44	.40	.07
45*	.33	.58
46	1.00	.00
47*	.47	.32
48	.65	-.16
49	.25	.23
50*	.79	.11
51	.93	.29
52*	.75	.36
53*	.70	.44
54*	.65	.21

Table 2 (Continued)

Item	Difficulty Level	Biserial Correlation Coefficient
55*	.44	.42
56	.93	.46
57*	.88	.22
58*	.77	.45
59*	.58	.17
60*	.58	.32
61	.12	-.21
62	.26	.27
63*	.70	.29
64*	.65	.16
65	.09	.63
66	.95	.31
67*	.86	.56
68	.96	.11
69	.95	.20
70	.67	.05
71	.26	.39
72*	.77	.31
73*	.86	.27
74*	.51	.46
75	.95	.71
76	.42	-.09
77*	.58	.23
78*	.77	.49
79*	.40	.11
80*	.86	.21
81*	.56	.41
82*	.89	.19
83*	.65	.58

Table 2 (Continued)

Item	Difficulty Level	Biserial Correlation Coefficient
84	.82	.00
85*	.67	.45
86*	.84	.39

*Indicates items retained for revised instrument.

.75. The usefulness of the instrument would be improved if the reliability coefficient were .80 or higher because a reliable coefficient of this magnitude is adequate for interpreting individual test scores (Nunnally, 1972, p. 106; Anastasi, 1976, p. 109).

Items were analyzed quantitatively in terms of their statistical properties, and qualitatively in terms of their content and form. Quantitative analysis principally included the measures of item difficulty and biserial correlation. Item difficulty is the proportion of correct answers on the item. The easier the item the higher the value. The biserial correlation is an estimate of the extent to which the item discriminates between examinees with high scores and examinees with low scores by using the total test score as a criterion.

Employing the criteria established by the test author for item analysis, items with a biserial correlation coefficient of .30 or above that ranged in difficulty from .31 through .91 were judged acceptable. Thirty-four items were subsequently identified that met this criterion. Because the reliability and content validity of the instrument could be jeopardized with a test of this length, the author decided to retain items with biserial correlation coefficients above .10. In applying the criterion, 58 items were retained for inclusion in the revised form of the instrument and 28 items were eliminated (see Table 3). The items that were eliminated exhibited one or more of the following characteristics: (1) very easy, (2) very difficult, or (3) poor discriminator. In addition,

Table 3

Items Grouped by Biserial Correlation Coefficients and Difficulty Levels

Field Tryout I

	Biserial Correlation Coefficients						
Difficulty	-.00 - .00	.01 - .09	.10 - .19	.20 - .29	.30 - .39	.40 & above	Totals
.92 - 1.00	46	7, 22	68	12, 17, 51, 69	66	4, 56, 75	12
.31 - .91	13, 48, 84, 76	44, 70	18, 26, 50, 59, 64, 79, 82	2, 3, 19, 20, 24, 30, 34, 35, 36, 41, 43, 54, 57, 63, 73 77, 80	10, 11, 16, 21, 25, 28, 38, 47, 52, 60, 72, 86	1, 5, 6, 8, 27, 29, 31, 32, 33, 37, 39, 40 45, 53, 55, 58, 67, 74, 78, 81, 83, 85	64
.00 - .30	15, 61	14	23	49, 62	42, 71	9, 65	10
Totals	7	5	9	23	15	27	86

eleven of the deleted items had a distractor chosen by more examinees than the designated answer (items 9, 14, 15, 23, 42, 44, 49, 61, 62, 65, and 71). The source of difficulty seemed to be that several answers could be the best answer, depending upon one's frame of reference.

Revising the Preliminary Test Form

Items which discriminated poorly (items 18, 26, 50, 59, 64, 79, and 82) were studied carefully to try to determine the source of the problem and were revised. Items with discrimination indices of .30 or above were also examined to determine if revision might improve their discriminating power and were subsequently revised if it seemed appropriate to do so. All distractors that were nonfunctioning were either replaced or reworded in an effort to make them more attractive. Finally, two items were written for Category I because only one item testing knowledge of objectives had survived the first field tryout. The revised instrument had 16 items in Category I, 31 items in Category II, and 13 items in Category III (see Appendix G). Finally, both the test directions and personal data sheet were simplified. The definition of subskills approach was eliminated in the test directions because qualifying phrases in item stems provided the necessary clarification. Revisions in the personal data sheet included deleting requests for: (1) number of hours of reading coursework, (2) school organization, and (3) specification of the number of years employed in reading related positions.

Reassessing Content Validity

The revised instrument was then submitted to expert review to determine if the test items measured the content outlined in the test base.

To facilitate the evaluation, each judge was given: (1) a copy of the items arranged in categories that corresponded to those in the test outline, and (2) a copy of the test outline (see Appendix H).

The four reviewers who had judged the preliminary set of items and five other persons who possessed similar educational backgrounds and professional experience concurred that the items appeared to measure the three categories in the test outline. After this final determination of content validity the items and the directions were assembled into the second field tryout form (see Appendix I).

Field Tryout II

The fourth phase of instrument development included: (1) securing teaching for the second field tryout of the instrument, (2) administering the instrument to the teachers, and (3) analyzing the resulting data.

Examinees

As in the first field tryout, the examinees for Field Tryout II were teachers in grades one through six who were selected on the basis of availability and willingness to participate. The sample

consisted of 104 teachers in four school districts in Westcentral and Southern Wisconsin. The pupil enrollments by district for the 1978-79 school year were: (1) 764, (2) 2,789, (3) 10,460, and (4) 13,471. Demographic data were collected for each examinee along with the test responses. These data were gathered in order to describe the population and to use as a classificatory means to: (1) divide the sample into two comparison groups on the basis of number of reading courses, and (2) examine the correlations between number of reading courses and years of teaching experience to test scores. Teachers were requested to provide the following information: (1) number of reading courses, (2) number of years of teaching experience, (3) grade level at which most of their teaching was done, and (4) whether or not they use or have used a skills management system.

Table 4 presents the number of reading courses completed by the teachers. Twenty percent of the teachers reported they had no reading course or one reading course, 25 percent had two reading courses, and 55 percent had 3 to 10 reading courses. The mean number of reading courses was 3, the median was 3, and the mode was 2. Data on the years of teaching experience are presented in Table 5. Although 59 of the teachers had taught 10 years or less, the mean years of teaching experience was 11 years. The sample consisted of 46 teachers in the primary grades and 58 teachers in the intermediate grades. Fifty-two percent of the primary level teachers reported using a skills management system while 50 percent

Table 4

Number of Reading Courses

Reported by Second Field Tryout Sample

Level	Number of Reading Courses											
	0	1	2	3	4	5	6	7	8	9	10	Total
Primary Teachers	1	7	10	10	7	2	5	1	2	1	0	157
Intermediate Teachers	1	12	16	10	7	6	2	2	1	0	1	176
TOTAL	2	19	26	20	14	8	7	3	3	1	1	333

Table 5

Years of Elementary Teaching Experience

of Second Field Tryout Sample

Years of Experience	Number of Teachers	Years of Experience	Number of Teachers
1	5	15	4
2	5	16	2
3	4	17	2
4	5	18	2
5	10	20	3
6	6	22	2
7	6	23	1
8	7	24	1
9	3	25	2
10	8	27	2
11	5	30	1
12	3	35	2
13	5	40	1
14	6		

of the intermediate level teachers reported using a skills management system.

Administration

The revised instrument was administered to groups of teachers by the author or one of the examiners who had proctored sessions in the first field tryout. Although no time limits were imposed, all examinees completed the test in 30 to 50 minutes. As before, an attempt was made to minimize the score variance that results from factors external to the examinees. This was done by: (1) securing rooms for the testing sessions that were free from distracting noises, (2) following uniform administration procedures parallel to those in the first field tryout (see Appendix J), and (3) allowing examinees to complete the test without interruptions.

Analysis of Data

The data from the second field tryout of the instrument was statistically analyzed with the GITAP program to determine: (1) the summary statistics for the total sample, (2) the Hoyt internal consistency reliability coefficient of the revised instrument, and (3) the difficulty level and biserial correlation coefficient of each item.

To find whether scores on the instrument reflected differences greater than the .05 level of significance between teachers with no reading course or one reading course and those with three or more

reading courses, the two-sample one-tail t-test was applied to the appropriate group mean scores. Pearson product-moment correlation coefficients were computed to assess the degree of relationship between: (1) number of reading courses and performance on the test, and (2) years of teaching experience and performance on the test.

Data from the second field tryout, which provides evidence of reliability and content validity for the revised instrument are presented and examined in Chapter IV.

CHAPTER III

PRESENTATION, ANALYSIS AND DISCUSSION OF DATA

In this chapter, data relevant to the performance of the revised instrument in the second field tryout are presented and examined. The order of presentation is: (1) reliability of the instrument, (2) content validity of the instrument, and (3) demographic data of the sample.

Evidence of Reliability

The data from the second field tryout of the instrument were analyzed with the GITAP program to assess the reliability of the revised instrument (see Table 6). Alternate ways of expressing reliability were provided by the reliability coefficient, which estimated the relative precision of the test for the particular group of examinees, and by the standard error of measurement, which provided an indication of the absolute accuracy of the test scores.

The reliability coefficient computed with the Hoyt internal consistency procedure was .82 and the standard error of measurement for the test group was 3.29. Thus, the reliability coefficient had improved substantially from a .75 coefficient obtained for the preliminary test form. Because a reliability coefficient in the .80's is generally considered to be adequate for interpreting individual measurement (Furst, 1958, p. 323; Nunnally, 1972, p. 99; Anastasi, 1976, p. 109) one can place a fairly high degree of confidence in

Table 6

Difficulty Levels and Biserial Correlation Coefficients

of Items in Field Tryout II

Item	No. on Field Tryout I	Difficulty Level	Biserial Correlation Coefficient
1.	(39)	.51	.46
2.	(3)	.88	.55
3.	(27)	.78	.35
4.	(67)	.77	.35
5.	(55)	.57	.34
6.	(60)	.53	.28
7.	(43)	.53	.43
8.	(80)	.60	.27
9.	----	.63	.28
10.	(40)	.63	.67
11.	(30)	.70	.18
12.	(36)	.72	.14
13.	(81)	.78	.47
14.	(37)	.84	.37
15.	(6)	.53	.41
16.	(29)	.61	.30
17.	(11)	.78	.52
18.	(82)	.93	.79
19.	(10)	.55	.60
20.	(85)	.68	.17
21.	(41)	.88	.49
22.	(34)	.73	.49
23.	(25)	.62	-.00
24.	(45)	.57	.38
25.	(53)	.80	.34
26.	(20)	.89	.73

Table 6 (Continued)

Item	No. on Field Tryout I	Difficulty Level	Biserial Correlation Coefficient
27.	(38)	.82	.70
28.	(21)	.92	.45
29.	(78)	.68	.25
30.	(32)	.75	.57
31.	(77)	.95	.62
32.	(35)	.62	.51
33.	(8)	.79	.55
34.	(1)	.54	.55
35.	(50)	.67	.27
36.	(74)	.45	.48
37.	(24)	.43	.25
38.	(54)	.91	.34
39.	(16)	.55	.42
40.	(31)	.28	.05
41.	----	.59	.42
42.	(26)	.86	.41
43.	(19)	.36	.18
44.	(57)	.88	.61
45.	(58)	.83	.57
46.	(33)	.55	.55
47.	(52)	.61	.55
48.	(86)	.87	.67
49.	(59)	.23	.28
50.	(5)	.54	.65
51.	(18)	.41	-.02
52.	(83)	.59	.36
53.	(2)	.57	.35
54.	(28)	.76	.59
55.	(72)	.75	.50

Table 6 (Continued)

Item	No. on Field Tryout I	Difficulty Level	Biserial Correlation Coefficient
56.	(64)	.73	.47
57.	(63)	.78	.29
58.	(47)	.58	.25
59.	(79)	.74	.28
60.	(73)	.63	.54

the consistency with which the test measures the performance of similar groups. The standard error of measurement is small, which indicates that two-thirds of the time a teacher's true score lies within a range of 3.29 points above and below the obtained score on the instrument. Because the reliability coefficient is fairly high and the standard error of measurement is small, the revised instrument is not dominated by measurement error and one can place confidence in the precision or consistency of performance of the instrument.

While test reliability is expressed statistically with the reliability coefficient and the standard error of measurement, additional information that is necessary to interpret the reliability estimate includes the characteristics of the group to which the instrument was administered, the method of its administration, and item analysis data.

Characteristics of the examinee group which are indicators of instrument reliability include the number of examinees, the range and mean of raw scores, and the standard deviation. The second field tryout group consisted of 104 elementary teachers which is adequate for stability of the reliability estimate with the Hoyt procedure. Scores of the test group ranged from 18 to 55 with a mean score of 40, a median of 41, and a mode of 43. The standard deviation was 7.76. Ebel (1972, p. 378) states that a standard deviation of one-sixth the range between the highest possible score and the expected chance score is quite satisfactory. This value is

7.50 for the instrument. Because the standard deviation of the set of test scores is satisfactory, the items tend to discriminate clearly between examinees at different ability levels. If the standard deviation for the test had been small this would have been an indication that the test was too easy, too hard, or composed of too many poorly discriminating items. Since the average score is near the ideal mean score of 37.5 and the standard deviation closely approximates 7.50 the test is efficient.

Because reliability is also a function of the uniformity of testing conditions, the author attempted to control factors that could contribute to instability in test performance. First, rooms were reserved for the testing sessions that were free from noise. Second, examinees completed the test without interruptions or time limits. Since the time required for completion of the test was generally 40 minutes, fatigue was not considered to be a factor influencing test reliability. Third, test examiners followed uniform procedures for administering the test which paralleled those in the first field tryout.

Further information about the instrument's internal consistency was provided when item difficulty and the biserial correlation coefficient were used as a pair to look at each item (see Table 7). Employing the criteria established by the author for item analysis, items that were in a midrange of difficulty (.40 - .90) and that had a biserial correlation coefficient of .30 or higher were identified as maximum discriminating items. Thirty-nine of the 60

Table 7

Items Grouped by Biserial Correlation Coefficients and Difficulty Levels

Field Tryout II

	Biserial Correlation Coefficients						
Difficulty	-.00 - .00	.01 - .09	.10 - .19	.20 - .29	.30 - .39	.40 & above	Totals
.91 - 1.00					38	18, 28, 31	4
.40 - .90	23, 51		11, 12 20	6, 8, 9, 29, 35, 37, 57, 58, 59	3, 4, 5, 14, 16, 24, 25 52, 53	1, 2, 7, 10, 13, 15 17, 19, 21, 22, 26, 27, 30, 32, 33, 34, 36, 39, 41, 42, 44, 45, 46, 47, 48, 50, 54, 55, 56, 60	53
.00 - .39		40	43	49			3
Totals	2	1	4	10	10	33	60

items met these criteria while only three items did not meet either of these criteria (items 40, 43, and 49). Of the 13 additional items that were in the .40 - .90 difficulty range, nine had biserial correlation coefficients of .20 - .29 (items 6, 8, 9, 29, 35, 37, 57, 58, and 59) while five had discriminating indices of less than .20 (items 11, 12, 20, 23, and 51). Although four other items discriminated adequately (.30 or above) they were comparatively easy (items 18, 28, 31, and 38). Further analysis indicated that in four of the items examinees selected a distractor rather than the keyed option (items 40, 43, 49 and 51).[2] In summary, by eliminating the 28 items on the preliminary test form that fell outside the midrange of difficulty and by increasing the discriminating power of 37 of the 60 items, test reliability increased to the point that the present instrument can be considered adequate for interpreting individual scores. Finally, the Hoyt internal consistency reliability of .82 appeared to compare favorably with test reliability reported for the instruments reviewed.

---

[2]Item 40 *b(a), Item 43 *d(c), Item 49 *b(a, c), Item 51 *d(c)
* = keyed option; ( ) = distractor(s) chosen.

Evidence of Validity

Because content validity is foundational in establishing the validity of achievement and proficiency tests, it seemed the most crucial type of validity for the present instrument. Content validity was built into the instrument from the very beginning of the test development process, primarily through systematic, logical analysis. It essentially involved the judgment of the author and subject matter experts to determine whether the outline of the test content was adequately specified and if all major aspects were measured by the test items. The specific procedures used to establish content validity were:

1. Evaluation of the outline of test content.
2. Initial review of the items.
3. Second review of the items.
4. Reassessment of content validity.

The descriptive outline of test content, developed from a review of textbooks and articles relating to a subskills approach, was submitted for evaluation to four persons who were qualified to judge the adequacy of the test outline. They were asked to judge whether the three content categories were appropriately defined and emphasized. Responses of the evaluators indicated that, in their opinion, the content categories were appropriate, both in definition and emphasis.

The author then developed 100 items to measure the content

outlined in the test base. This phase in the test development process involved making a number of decisions. First, it was necessary to derive a set of topics for items from the outline of test content. Second, content that would be translated into items was identified by careful study of a given set of materials that are used in preservice and in-service courses. The determination of content that was most relevant was based on whether it: (1) sampled the three categories, and (2) represented a fair degree of consensus among authorities.

After the items were generated, they were submitted to four subject matter experts (two had measurement backgrounds) for review. Reviewers responded to the following questions for each item:

1. Is the item keyed to the appropriate category?
2. Do you agree that the best answer is the best answer?

The extent of agreement among the judges, determined with Cronbach's Alpha, was .85 for the first question and .91 for the second question. In a joint meeting, reviewers identified items that should be revised and/or replaced and recommended ways to improve the items. They agreed that items testing knowledge of comprehension skills needed to be replaced and one reviewer felt more items on transfer should be written. Then, based on the comments of the reviewers, items were revised or deleted and replaced.

A further check on content validity was made when the items were keyed independently by 12 persons. After their responses to each of the items were compiled, seven of the persons were

interviewed and acted as consultants by furnishing information on the items. Based on their comments, further revisions were made in the instrument.

Following the first field tryout, 28 items were deleted, a number of the remaining items were extensively revised, and two new items were written to test knowledge of behavioral objectives. As a final check of content validity the revised instrument was submitted to expert reviewers to reevaluate the adequacy of the items for sampling the content categories. The group of nine judges, which included the initial reviewers, agreed that the items appeared to measure the three categories in the test outline.

The cumulative effect of decisions made during the test development process resulted in allotting proportionally more items to Category II than to the other two categories. Although the categories could have been equalized by writing more items for Categories I and III or by randomly deleting a number of items in Category II, neither of these options was chosen because of the probable effect on reliability and content validity. Because the test would be more useful if the essential aspects of Category II were sampled, items that survived item analysis were retained for the revised instrument. The author concluded that the distribution of items, as a whole, covered what the test was intended to cover and represented the outline of test content even though equal emphasis was not observed. In summary, it appeared that adequate evidence existed for judging the content validity of the revised instrument.

Huddleston (1956) states that for a test to eventually demonstrate content validity, it "should be valid at the planning stage, the writing stage, the reviewing stage, and at the assembly stage" (p. 283). At each step of the test development process, expert judgment and logical analysis determined the adequacy of individual items so that, in the final analysis, the instrument seemed to serve the purpose for which it was intended. Finally, a comparison of the evidence for judging content validity reported for the tests reviewed appeared to compare favorably with evidence provided by the author.

Demographic Data

Demographic data related to experience and training were supplied by examinees in the second field tryout group. The sample consisted of 46 primary grade teachers and 58 intermediate grade teachers. Years of teaching experience ranged from 1 to 40 years. Although 59 teachers had taught 10 years or less, the mean years of teaching experience was 11 years. Training in reading courses ranged from 0 to 10, with the mean and median number of reading courses 3, and the mode 2.

If the data for the two field tryout samples are compared (see Table 8), it can be noted that the samples in both field tryouts included a greater number of intermediate grade teachers than primary grade teachers. In terms of number of reading courses, the two samples were comparable. The range of years of teaching experience was also similar although the mean years was 11 for the second

Table 8

Descriptions on Variables of Experience and Number of Reading Courses for Samples in Field Tryouts of the Test

Sample	Variables of Experience and Reading Courses							
	Teaching Level		Number of Reading Courses				Years of Experience	
	Primary	Intermediate	Range	Mean	Median	Mode	Range	Mean
Field Tryout I								
N = 57	22	35	0-10	3.63	3	2	1-37	9.56
Field Tryout II								
N = 104	46	58	0-10	3.00	3	2	1-40	11.00

tryout sample and 9.56 for the first tryout sample.

Following the test administration, examinees were divided into two comparison groups: (1) teachers with zero or one reading course, characterized as novice, and (2) teachers with three or more reading courses, characterized as expert. The novice group consisted of 21 teachers and the expert group consisted of 57 teachers. A one-tail t-test was used to determine whether differences between the group mean scores were significant at the .05 level. The resulting separate variance estimate indicated a t-value of -.31, less than the critical value of t (1.98) required for significance at the .05 level (see Table 9). A possible explanation for the lack of difference between the two sample means is that what individuals know about a subskills approach is learned in the first reading course. Another possibility is that since skill-centered teaching is a fairly recent phenomenon, teachers with more reading courses may have acquired their reading knowledge before a subskills approach was emphasized in reading coursework. Or, perhaps the instrument failed to differentiate the hierarchy of knowledge to be acquired with increasing preparation in reading. It is possible that the variable doesn't account for variability in the scores because what is sampled is so basic or global it is common knowledge. Taking reading courses may not have anything to do with the test score because the instrument may not be measuring content covered in courses. Or, perhaps the test is not measuring the degree of knowledge individuals have about a subskills approach.

Table 9

Comparison of the Difference Between Sub-Group Mean Scores for Number of Reading Courses of Second Tryout Sample

No. of Reading Courses	No. of Teachers	Mean	S.D.	t	p
0 - 1	21	40.29	5.29		
				-.31	NS
3 or more	57	40.77	8.01		

The following relationships were also examined: (1) number of reading courses and performance on the instrument, and (2) years of teaching experience and performance on the instrument. The Pearson product-moment correlation between number of reading courses and performance on the instrument was computed at .14, which indicates little relationship between getting a high score and a large number of reading courses. The Pearson product-moment correlation between years of teaching experience and performance on the instrument was -.31, which indicates a statistically significant negative relationship, at the .001 level, between score and years of teaching experience. Although a negative correlation indicates that a decline in performance is associated with increasing age and increasing time out in the field, one can not conclude that experience causes the decline. They may be correlated because they are both related to a third variable. One could speculate that: (1) since a subskills approach is a fairly recent phenomenon, older teachers may have acquired their knowledge of reading instruction before skill-centered teaching was emphasized, (2) older teachers who are knowledgeable may have dropped out or moved to specialist positions, (3) teachers may become restricted to implementing a prescribed program of instruction, and/or (4) motivational variables on the part of the teacher may be contributing to the relationship. Nevertheless, as teacher experience increases knowledge of skill-centered teaching decreases. Finally, based on the reports of the test authors who investigated the relationship of training in reading and

teaching experience to test scores, it has not been clearly demonstrated that either variable is consistently and positively related to performance on a knowledge test. In terms of reading training, five authors found reading knowledge to be a positive factor (Burnett, 1961; Broman, 1962; Thompson, 1972; Artley & Hardin, 1975; Thomas, 1975) while four authors found a reading background to have no apparent effect (Shannon, 1959; Farinella, 1960; McCollum, 1964; and Henriksen, 1968). In terms of teaching experience, six authors (Aaron, 1960; Wade, 1960; Burnett, 1961; Broman, 1962; Spache & Baggett, 1965; Thomas, 1975) found a positive effect while four authors (Shannon, 1959; Henriksen, 1967; McCollum, 1972; Thompson, 1972) found no apparent effect.

Summary

In summary, data relevant to the performance of the present instrument were presented and examined. Because a reliability coefficient of .82 and a standard error of measurement of 3.29 was obtained for the instrument, one can place confidence in the consistency with which the instrument measures the performance of similar groups. Content validity was built into the instrument from the beginning. At each step of the test development process, expert judgment and logical analysis determined the adequacy of the items for measuring the test outline. In the final analysis, the instrument seems to serve the purpose for which it was intended.

The summary, conclusions, implications and recommendations are presented in Chapter IV.

CHAPTER IV

SUMMARY, CONCLUSIONS, LIMITATIONS, AND RECOMMENDATIONS

This chapter represents a summary of the procedures followed in construction of an instrument to assess teachers' knowledge of a subskills approach to teaching reading. Included in the chapter are the conclusions, limitations, and recommendations which appear to be warranted on the basis of the data reported.

Summary of the Research

This is a report of the procedures followed in constructing and field testing an instrument to assess elementary teachers' knowledge of a subskills approach to teaching reading. The rationale for assessing knowledge of skill-centered teaching was based on several factors. First, the available assessment instruments are unfocused and, consequently, knowledge of the content of reading instruction has not been systematically examined. Second, knowledge that directly relates to the teaching of reading has not been specifically investigated. Finally, basic to the significance of development of this instrument was the assumption that a skills orientation to reading instruction is pedagogically efficient and effective and, therefore, provides a foundation for competency in teaching reading.

In the first phase of instrument development an outline of test content was developed and was submitted to four qualified judges for evaluation. The second phase involved writing and selecting items for the preliminary test form. Before item writing began, specific

knowledge competencies were derived from the test outline and content that would be translated into items was identified. A pool of multiple choice items was subsequently developed and submitted to a critical review of persons knowledgeable in the subject matter area. Based on the reviewer's comments, items were revised and/or replaced with new items. A further check of content validity was made when knowledgeable persons keyed the items and acted as consultants by furnishing information for improving the preliminary test form. The third phase of the test construction process involved: (1) field testing the preliminary test form with elementary teachers, (2) analyzing and revising the items based on the field tryout data, and (3) reassessing content validity of the revised instrument. In the fourth phase of test construction, the revised instrument was subjected to a second field tryout with 104 elementary teachers. The resulting data were examined to determine reliability and content validity of the present instrument. In addition, a one-tail t-test was used to determine whether differences between the group mean scores for teachers with zero or one reading course and teachers with three or more reading courses were significant at the .05 level. The resulting separate variance estimate indicated a t-value of -.31, less than the critical value of t (1.98) required for significance at the .05 level. Finally, the correlations between number of reading courses and years of teaching experience to test scores were determined. The Pearson product-moment correlation between number of reading courses and performance on the instrument was .14, which

indicates little relationship between getting a high score and a large number of reading courses. The Pearson product-moment correlation between years of teaching experience and performance on the instrument was -.31, which indicates a statistically significant negative relationship, at the .001 level, between score and years of teaching experience.

Conclusions

An examination of the data from the second field tryout indicated that the present instrument, with a Hoyt internal consistency reliability coefficient of .82 and a standard error of measurement of 3.29, may be considered adequate for interpreting individual test scores. Because the reliability coefficient is fairly high and the standard error of measurement is small, 3.29, the instrument is not dominated by measurement error and one can place confidence in the consistency with which the test measures the performance of individuals similar to those in the field tryout sample. Additional statistical indicators of test reliability included: (1) a range of raw scores from 18 to 55, (2) a mean score of 40, (3) a standard deviation of 7.76, and (4) item analysis data which identified 39 of the 60 items as maximum discriminating items. Based on this evidence of the present instrument's internal consistency, one can conclude that the test is efficient because the items are not too easy or too hard and they tend to discriminate clearly between examinees at different ability levels.

A review of the test development process suggests that the present instrument demonstrates content validity. The basis for this conclusion is that expert judgment of the adequacy of the items is reflected at each step of instrument development. The content validity of the instrument was derived, in the planning stage, from a specification of the test outline and agreement as to its adequacy by four qualified judges. During the item writing stage, further evidence of content validity was provided by generating items which measured knowledge competencies in line with the test outline. Content that was translated into specific items was identified in a careful study of materials that are representative of those teachers encountered in their pre-service and in-service coursework. To achieve content validity during the reviewing stage, an initial review of the items and a subsequent tryout review by knowledgeable persons preceded the first field tryout of the preliminary test form. Finally, a reassessment of content validity of the revised instrument, prior to the second field tryout, indicated that competent judges agreed that the items sampled the content categories stated in the test outline. Thus, based on the cumulative effect of systematic, logical analysis by the test author and the subject matter experts, content validity was built into the instrument from the beginning of the test development process.

Limitations

Any conclusions that seem tenable regarding the test must be tempered by the following limitations.

1. Examinees were not selected at random, as representative of all Wisconsin elementary teachers. Rather, examinees in both field tryouts of the instrument were elementary teachers selected on the basis of availability and willingness to participate in the research. In some instances, the total faculty in a school did not elect to take the test. The fact that all teachers did not take the test may have had some effect on reliability.
2. The parameters of the test based and the relevance of item content reflects the understandings, values, and limitations of a single test author.
3. This instrument was developed only the point of establishing content validity. Predictive validity remains to be established for the instrument.

Recommendations

This discussion is focused on several recommendations for future research on the instrument and utilization of the instrument which became apparent as the test construction process was completed. Because the instrument was carefully constructed it provides the

basis for assessing what teachers know about a subskills approach to teaching reading. First, it is possible to use the test to assess what individuals learn in a course that emphasizes skill-centered teaching. The instrument may be administered before and after instruction designed to acquaint students with the components of a subskills approach. Furthermore, the instrument can be used to focus the planning of in-service sessions related to reading. By analyzing teachers' responses to the test items, in-service leaders can identify gaps and limitations in teachers' knowledge and then plan sessions to provide the needed information. In essence, the instrument may be utilized to assess both pre-service and in-service teachers' knowledge of a subskills approach to teaching reading.

The relationship of elementary teachers' experience in the field and performance on the instrument could also be examined further. The observation of a statistically significant inverse relationship between teachers' scores on the test and number of years of teaching experience could have several explanations. While a decline in test performance with increasing age and years of teaching experience is indicated, one can not infer that experience causes the decline. One could speculate that: (1) older teachers may have acquired their knowledge of reading instruction before skill-centered teaching was emphasized, (2) older teachers who are knowledgeable may have dropped out or moved to specialist positions, or (3) teachers become restricted to implementing a prescribed program of instruction so they forget what they had learned

relative to a subskills approach.

The first step in examining the inverse relationship would be to identify a large sample and to collect additional data in order to avoid a sampling bias. Then the test would be administered to see if the negative correlation would again be demonstrated. At the same time one could also obtain permission to examine transcripts of the examinees to determine how many reading courses they had taken, how recently they had completed the last course, and their over-all grade average. One would also document examinees' length of experience in the teaching of reading. Correlations could be computed to examine the relationship between test scores and each of these variables. Such information would allow one to understand more fully the relation of amount of training, recency of training, and academic performance level to test performance. Furthermore, by examining the records one could determine the actual course work completed by individuals and thereby correct for any errors in teachers' self reports. If there is a reliable decline in the test scores associated with increasing number of years since examinees had their last reading course, the cause could be a recency of training effect. If that were so, it would be reasonable to expect that individuals involved in continuing education would earn higher test scores. Finally, if there were a positive relationship between grade average and test performance one might next examine the relationship of measured intelligence to test performance.

A second recommendation is to determine if the instrument

differentiates between undergraduates who have no reading courses and undergraduates with a reading course that emphasizes a subskills approach to teaching reading. Given instruction in the content of the test outline, one would predict that undergraduates who had been exposed to instruction emphasizing components of a subskills approach would attain higher scores on the test than undergraduates who had not been exposed to this instruction. One could use either a one group pre-post-design or an after-only-design with a control group (Fox, 1969, pp. 473-4). With the one group pre-post-design, the test would be administered prior to instruction and again after instruction. With the after-only design with one control group, the test would be administered at the end of the instructional period to the group that received instruction and to a control group that received no instruction. Because in the latter instance data are collected only at the end of instruction, one must assume comparability of the groups at the beginning of the study. In both situations, the critical thing is to have control of what is taught. Then, if individuals are taught the content of the test outline there should be an increase in knowledge. If such a study were conducted the predictive validity of the instrument could be determined.

A suggestion for practical use of the instrument involves using it as a vehicle for staff development. The instrument would be administered to practicing teachers to find out what they need to learn about a subskills approach to teaching reading. Once the test results were analyzed to determine what teachers do not know,

then in-service programs could be planned to teach that particular content. Providing in-service programs that focus on topics identified in a systematic way promises to be far more worthwhile than providing unfocused sessions that reflect only teachers' or administrators' interests. One problem in using teacher-identified topics to shape in-service programs is that teachers may be unaware of what it is they need to learn about the teaching of reading. For example, because pre-service training may be inadequate in relation to helping teachers provide effective skill instruction for individual students, teachers may be unaware of the components of a subskills approach. Consequently, their suggestions for in-service programs would not be likely to include topics in this area, so any problems related to teaching skills would not be identified. By using the instrument as a diagnostic tool its usefulness would be extended. Thus, rather than merely labelling teachers as knowledgeable or not knowledgeable, one could use the instrument to focus in-service efforts and, presumably, to increase students' achievement.

Finally, further research on the instrument could focus on determining whether observations of teacher performance in the field are correlated with performance on the instrument. The test could be administered to in-service teachers or pre-service teachers. Behavioral criteria based on the content of the present instrument could be developed as a basis for an observational tool that would enable one to determine whether there is a relationship between teachers' knowledge and the extent to which they implement effective

skill-centered instruction. One would expect teachers who score high on the test to behave differently when they teach reading than individuals who score low. Thus, a criterion-related study would be a test of the instrument's utility in predicting actual behaviors.

REFERENCES

Aaron, I. E. What teachers and prospective teachers know about phonic generalizations. Journal of Educational Research, 1960, 53, 323-330.

Aaron, R. Research report summary. PLAN, an individualized learning system. Iowa City, Iowa: Westinghouse Learning Corporation, 1974.

Allen, J. E. The right to read -- target for the 70's. Elementary English, 1970, 47, 487-492.

Anastasi, A. Psychological testing (3rd ed.). New York: Macmillan Publishing Co., 1968.

Anastasi, A. Psychological testing (4th ed.). New York: Macmillan Publishing Co., 1976.

Artley, A. S., & Hardin, V. B. Inventory of teacher knowledge of reading. Columbia, Missouri: Lucas Brothers Publishers, 1975.

Askov, E. N., Otto, W., & Fischbach, T. Development of specific reading skills in adult education. NRC Conference Yearbook, 1971, 20, 46-51.

Athey, I. Essential skills and skills hierarchies in reading comprehension and decoding instruction. Paper presented at conference of the N.I.E., Washington, D. C., 1974.

Athey, I. (Chairperson). Panel 10: Essential skills and skill hierarchies in reading. N.I.E. Conference Report on Studies in Reading, Washington, D. C., 1975.

Athey, I. Comments on design for developing comprehension. In J. Guthrie (Ed.), Cognition, curriculum and comprehension. Newark: International Reading Association, 1977.

Bailey, M. The utility of phonic generalizations in grades one through six. The Reading Teacher, 1967, 20, 413-418.

Baker, F. B. Fortap: A fortran test analysis package. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, 1969.

Berliner, D. & Rosenshine, B. Classroom learning. In R. C. Anderson, R. J. Spiro, and W. E. Montague (Eds.), Schooling and the acquisition of knowledge. New Jersey: Lawrence Erlbaum Associates, Publishers, 1977.

Berry, B. T. Improving freshman reading habits. English Journal, 1931, 20, 824-828.

Blohm, P. J. Wisconsin design word attack skills and overall reading achievement. University of Wisconsin: Unpublished Monograph, 1975.

Blohm, P. J. Effects of skill-based instruction and skill mastery testing on general reading comprehension of tenth-grade poor comprehenders. Technical Report No. 449 (Doctoral Dissertation), Wisconsin Research and Development Center for Cognitive Learning, 1978).

Bloom, B. S. Taxonomy of educational objectives, handbook I, cognitive domain. New York: David McKay, 1956.

Bond, G. L. & Dykstra, R. The cooperative research program in first-grade reading instruction. Reading Research Quarterly, Summer 1967, 2, 5-142.

Broman, B. L. Factors associated with teacher knowledge of reading skills. Doctoral Dissertation, University of Tennessee, 1962. Dissertation Abstracts International, 1962, 23, 62-5574, 133.

Burmeister, L. E. Usefulness of phonic generalizations. The Reading Teacher, 1968, 21, 349-356, 360.

Burnett, R. W. The diagnostic problem solving proficiency of elementary school teachers in teaching reading. Doctoral Dissertation, University of Indiana, 1961. Dissertation Abstracts International, 1961, 22, 62-178, 98.

Carroll, J. B. The nature of the reading process. In H. Singer and R. B. Ruddell (Eds.), Theoretical models and processes of reading (2nd ed.), Newark: International Reading Association, 1976.

Chall, J. S. Learning to read: The great debate. New York: McGraw-Hill, 1967.

Chester, R. D. The comprehension component of the Design. NRC Conference Yearbook, 1972, 21, 258-264.

Clymer, T. The utility of phonic generalizations in the primary grades. The Reading Teacher, 1963, 16, 252-258.

Cohn, S. A. Reading for meaning: The reader's response. London: Hutchinson Educational Ltd., 1973.

Communication Skills Program. Englewood: Southwestern Regional Laboratory, 1971.

Conrad, H. S. The experimental tryout of test materials. In E. F. Lindquist (Ed.), Educational measurement. Washington, D. C.: American Council on Education, 1951.

Cronbach, L. J. Further evidence on response sets and test design. Educational and psychological measurement, 1950, 10, 3-31.

Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

Cronbach, L. J. Test validation. In Robert L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D. C., 1971.

Davis, F. B. Fundamental factors of comprehension in reading. Psychometrika, 1944, 9, 185-197.

Davis, F. B. Item selection techniques. In E. F. Lindquist (Ed.), Educational measurement. Washington, D. C., 1961.

Davis, F. B. Research in reading comprehension. Reading Research Quarterly, 1968, 3, 499-545.

Dewey, J. C. The acquisition of facts as a measure of reading comprehension. Elementary School Journal, 1935, 35, 346-348.

Duffy, G. G. Maintaining a balance in objective-based reading instruction. The Reading Teacher, 1978, 31, 519-523.

Duffy, G. G. & Sherman, G. B. Systematic reading instruction. New York: Harper & Row Publishers, 1977.

Durkin, D. Phonics test for teachers. New York: Teachers College Press, 1964.

Durkin, D. Fundamentals and principles underlying phonic instruction, reading and language. Proceedings of the International Reading Association, 1965, 10, 427-443.

Ebel, R. L. Measuring educational achievement. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1965.

Ebel, R. L. Essentials of educational measurement. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972.

Eisner, E. W. Instructional and expressive objectives: Their formulation and use in curriculum. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, 1969.

Eisner, E. W. Educational objectives -- help or hindrance? School Review, 1967, 75, 250-260.

Emans, R. The usefulness of phonic generalizations. The Reading Teacher, 1967, 20, 419-425.

Farinella, J. T. An appraisal of teacher knowledge of phonic analysis and structural analysis. Doctoral dissertation, University of Connecticut, 1960. Dissertation Abstracts International, 1960, 21, 60-5226, 204.

Feder, D. D. Comprehension maturity tests in a new technique in mental measurement. Journal of Educational Psychology, 1938, 29, 597-606.

Flip, J. Relationship among reading subskills: A hierarchical hypothesis. Journal of Reading Behavior, 1975, 7, 229-240.

Follman, J., Lowe, A. J., & Pfost, P. Reading comprehension: The empirical literature. NRC Yearbook, 1972, 22, 294-303.

Fox, D. J. The research process in education. New York: Holt, Rinehart and Winston, Inc., 1969.

Furst, E. J. Constructing evaluation instruments. New York: Longmans, Green, & Co., 1958.

Gagne, R. M. The acquisition of knowledge. Psychological Review, 1962, 69, 355-365.

Gibson, E. J. Learning to read. Science, 1963, 148, 1066-1072.

Gibson, E. J. & Levin, H. The psychology of reading. Cambridge: Massachusetts Institute of Technology, 1975.

Goodman, K. S. Reading: A psycholinguistic guessing game. Journal of the Reading Specialist, 1967, 4, 126-135.

Goodman, K. S. The reading process: Theory and practice. Paper presented at the annual meeting of the International Reading Association, Anaheim, California, May 1970.

Goodman, K. S. & Goodman, Y. Learning about psycholinguistic processes by analyzing oral reading. Harvard Educational Review, 1977, 47, 317-333.

Gray, W. S. The major aspects of reading. In Helen M. Robinson (Ed.), Sequential development of reading abilities, Supplementary Educational Monographs, No. 90. Chicago: University of Chicago Press, 1960, 8-24.

Green, D. & Roudabush, G. An investigation of bias in a criterion referenced reading test. Paper presented at the American Educational Research Association Meeting, Washington, D. C., March, 1975.

Gulliksen, H. Theory of mental tests. New York: John Wiley & Sons, Inc., 1950.

Guthrie, J. T. Models of reading and reading disability. Journal of Educational Psychology, 1973, 65, 9-18.

Harris, M. L. Some methodological suggestions for construction of an objective measurement instrument. Technical Memorandum, Wisconsin Research and Development Center for Cognitive Learning. The University of Wisconsin, 1968, No. M-1968-2.

Henriksen, E. B. An analysis of teacher knowledge of word recognition skills. Doctoral Dissertation, University of Chicago, 1967. Dissertation Abstracts International, 1967, 29, 69-9492.

Henrysson, S. Gathering, analyzing, and using data on test items. In Robert L. Thorndike (Ed.), Educational measurement, (2nd ed.). Washington, D. C., 1971.

Horn, M. They help me read new words: Inner city students learn reading by mastering basic subskills. CEDAR Report, 1974, 3, No. 1, 4-7. Washington, D. C.: U.S. Office of Education.

Hubbard, D. & Sals, D. The Wisconsin design for reading skill development; study skills -- A report on the type I field test, 1971-1973. Technical Report No. 341. Madison: Wisconsin Research and Development Center for Cognitive Learning, May, 1975.

Huddleston, E. M. Test development on the basis of content validity. Educational and Psychological Measurement, 1956, 16, 283-293.

Johnson, D. D. Word identification test for teachers. In R. Smith and D. Johnson, Teaching children to read. Reading, Massachusetts: Addison-Wesley Publishing Co., 1976.

Kingston, A. J., Brosier, G. F., & Hsu, Y. M. The inventory of teacher knowledge of reading -- a validation. The Reading Teacher, 1975, 29, 133-136.

Klausmeier, H. J. and Goodwin, W. Learning and human abilities: educational psychology (4th ed.). New York: Harper & Row Publishers, 1975.

Klopp, P. The Wisconsin design for reading skill development: comprehension -- A report on the 1973-1974 small scale field test. Technical Report No. 373. Madison: Wisconsin Research and Development Center for Cognitive Learning, 1976.

Lennon, R. T. What can be measured? The Reading Teacher, 1961-1962, 15, 326-327.

Mallett, J. J. The contribution of specific characteristics of first grade reading teachers concerning children's success in initial reading instruction. Doctoral Dissertation, University of Toledo, 1972. Dissertation Abstracts International, 1972, 33, 73-5940, 188.

Mason, J. M. A consideration of skill hierarchy approaches to the teaching of reading. Technical Report No. 42. University of Illinois: Center for the Study of Reading, 1977.

Mazurkiewicz, A. J. What do teachers know about phonics. Reading World, 1975, 14, 165-177.

McCollum, J. A. Teachers' knowledge of word analysis skills and linguistic concepts. Doctoral Dissertation, University of California, Berkeley, 1964. Dissertation Abstracts International, 1964, 25, 64-9051, 170.

Mehrens, W. A. & Lehmann, I. J. Measurement and evaluation in education and psychology. New York: Holt, Rinehart and Winston, Inc., 1973.

Morrison, C. & Austin, M. C. The torch lighters revisited -- A preliminary report. Reading World, 1976, 29, 647-652.

Nunnally, J. C. Educational measurement and evaluation. New York: McGraw-Hill Book Co., 1972.

Otto, W. Design for developing comprehension skills. In J. Guthrie, (Ed.), Cognition, curriculum and comprehension. Newark: International Reading Association, 1977.

Otto, W. & Askov, E. N. The Wisconsin design for reading skill development: Rationale and guidelines. Minneapolis: National Computer Systems, 1970.

Otto, W., Chester, R., McNeil, J., & Myers, S. Focused reading instruction. Reading, Massachusetts: Addison-Wesley Publishing Co., 1974.

Otto, W. & Chester, R. Objective-based reading. Reading, Massachusetts: Addison-Wesley Publishing Co., 1976.

Otto, W. & Harper, B. Constructing a test of teacher's knowledge of reading. Madison, Wisconsin: University of Wisconsin-Madison, 1977. (ERIC Document Reproduction Service No. ED 151 744).

Rafferty, P. T. Reading certification: The elementary teacher, English teacher, and reading specialist. M. S. in Education Thesis, Southern Illinois University, 1975, 60. (ERIC Document Reproduction Service No. ED 103 837).

Ramsey, W. Z. Will tomorrow's teachers know and teach phonics. The Reading Teacher, 1962, 15, 241-245.

Rorie, I. L. Analysis and validation of the inventory of teacher knowledge of reading. Doctoral Dissertation, University of Missouri, 1975. Dissertation Abstracts International, 1975, 36, 46-1041, 121.

Rosenkranz, C. Comprehension: A reading improvement project for EMR's. Gillett, Wisconsin: Cooperative Service Agency, No. 3. Bureau Memorandum, 1973, XIV, No. 4, 25-26.

Rosenshine, B. Academic engaged time, content covered, and direct instruction. University of Illinois: Unpublished manuscript, 1977.

Rystrom, R. Toward defining comprehension: A second report. Journal of Reading Behavior, 1970, 2, 157-158.

Samuels, S. J. Hierarchical subskills in the reading acquisition process. Invited address, Hyman Blumberge Symposium on Research in Early Childhood Education, John Hopkins University, November 1974.

Schubert, D. G. Teachers and word analysis skills. Journal of Developmental Reading, 1959, 2, 62-64.

Shannon, M. The measurement of phonic understandings relevant to the teaching of reading. Unpublished Ph.D. Dissertation, Harvard University, 1959.

Smith, F. Making sense of reading -- and of reading instruction. Harvard Educational Review, 1977, 47, 386-395.

Smith, F. Understanding reading (2nd ed.). New York: Holt, Rinehart, and Winston, Inc., 1978.

Smith, R. J., Otto, W., & Hansen. The school reading program. Boston: Houghton Mifflin Co., 1978.

Sochor, E. E. The nature of critical reading. Elementary English, 1959, 36, 47-58.

Soltis, J. F. An introduction to the analysis of educational concepts. Reading, Massachusetts: Addison-Wesley Publishing Co., 1968.

Spache, G. & Baggett, M. E. What do teachers know about phonics and syllabication. The Reading Teacher, 1965, 19, 96-99.

Spache, G. D. & Spache, E. B. Reading in the elementary school (4th ed.). Boston: Allyn and Bacon, 1977.

Spearritt, D. Identification of subskills of reading comprehension by maximum likelihood factor analysis. Research Bulletin RB-72-4. Princeton, New Jersey: Educational Testing Service, 1972.

Stallard, C. Comparing objective based reading programs. Journal of Reading, 1977, 21, 36-44.

Stennett, R. G., Smythe, P. C., & Hardy, M. Hierarchical organization of reading subskills: Statistical approaches. Journal of Reading Behavior, 1975, 7, 223-227.

Thomas, D. G. The construction of a criterion-referenced test to measure the ability of elementary school teachers to choose and interpret data for assessment in reading. Doctoral Dissertation, University of Kentucky, 1975. Dissertation Abstracts International, 1975, 36, 76-6149, 369.

Thompson, F. E. Teachers' knowledge of selected skills from the Florida catalog of reading objectives. Doctoral Dissertation, Florida State University, 1972. Dissertation Abstracts International, 1972, 33, 60-06071, 167.

Thorndike, R. L. Reading as reasoning: A study of mistakes in paragraph reading. Journal of Educational Psychology, 1917, 8, 823-832.

Thorndike, R. L. & Hagen, E. Measurement and evaluation in psychology and education (3rd ed.). New York: John Wiley and Sons, 1969.

Tinkelman, S. N. Planning the objective test. In Robert L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D. C., 1971.

Touton, F. C. & Berry, B. T. Reading comprehension at the junior high level. California Quarterly of Secondary Education, 1931, 6, 245-251.

Vanroosendaal, M. L. Selected factors involved in elementary teachers' knowledge of the teaching of reading. Doctoral Dissertation, Northern Illinois University, 1976. Dissertation Abstracts International, 1976, 36, 76-09867, 112.

Vaughn, K. W. Planning the objective test. In E. F. Lindquist (Ed.), Educational measurement. Washington, D. C.: American Council on Education, 1951.

Wade, E. W. The construction and validation of a test of ten teacher skills used in reading instruction. Doctoral Dissertation, Indiana University, 1960. Dissertation Abstracts International, 1960, 22, 60-6071, 91.

Wesman, A. G. Writing the test item. In Robert L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D. C., 1971.

APPENDIX A

EVALUATION FORM FOR TEST OUTLINE

Evaluation Form

1. Do you agree or disagree with respect to the three categories on the basis of their; (1) definitions, and (2) equal emphasis (equal number of items)?

		Agree	Disagree
Category I			
	Definition	____	____
	Emphasis	____	____
Category II			
	Definition	____	____
	Emphasis	____	____
Category III			
	Definition	____	____
	Emphasis	____	____

2. Are there other topics that have not been included? _____

   Please write a brief description of these topics.

   _____

   _____

   _____

   _____

3. Are there topics that should be deleted or revised? _____

   Please indicate those topics.

   _____

   _____

   _____

   _____

(CONTINUED ON PAGE 2)

4. If the weights are not appropriate, please indicate preferable weights.

_____

_____

_____

_____

Evaluator_____

Position_____

Date_____

APPENDIX B

SOURCE MATERIALS FOR ITEM CONTENT

Source Materials for Item Content

Duffy, G. G. & Sherman, G. B. Systematic reading instruction (2nd edition). New York: Harper & Row Publishers, 1977.

Durkin, D. Teaching them to read (2nd ed.). Boston: Allyn and Bacon, Inc., 1974.

Durkin, D. Teaching young children to read (2nd ed.). Boston: Allyn and Bacon, Inc., 1976.

Gagne, R. M. Essentials of learning for instruction. Hinsdale, Illinois: The Dryden Press, 1974.

Harris, L. A. and Smith, C. B. Reading instruction (2nd ed.). New York: Holt, Rinehart, and Winston, 1976.

Heilman, A. W. Principles and practices of teaching reading (4th ed.). Columbus: Charles E. Merrill Publishing Co., 1977.

Karlin, R. Teaching elementary reading (2nd ed.). New York: Harcourt Brace & Jovanovich, Inc., 1975.

Klausmeier, H. J. & Goodwin, W. Learning and human abilities: educational psychology (4th ed.). New York: Harper & Row Publishing Co., 1975.

Otto, W.; Chester, R.; McNeil, J.; and Myers, S. Focused reading instruction. Reading, Massachusetts: Addison-Wesley Publishing Co., 1974.

Pearson, P. D. and Johnson, D. D. Teaching reading comprehension. New York: Holt, Rinehart, and Winston, 1978.

Smith, R. J. and Johnson D. D. Teaching children to read. Reading, Massachusetts: Addison-Wesley Publishing Co., 1976.

Smith, R. J. and Barrett, T. C. Teaching reading in the middle grades. Reading, Massachusetts: Addison-Wesley Publishing Co., 1976.

Spache, G. D. and Spache, E. B. Reading in the elementary school (4th ed.). Boston: Allyn and Bacon, 1977.

Tinker, M. A. and McCullough, C. M. Teaching elementary reading (4th ed.). Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1975.

Zintz, M. V. The reading process (2nd ed.). Dubuque, Iowa: William C. Brown Co. Publishers, 1975.

APPENDIX C

INSTRUCTIONS AND EVALUATION FORM FOR ITEM REVIEW

AN EVALUATION OF ITEMS AND DIRECTIONS
FOR A TEST TO ASSESS KNOWLEDGE OF A
SUBSKILLS APPROACH TO TEACHING READING

Evaluator _____

Position _____

Date _____

To the reviewer: A preliminary draft of test items and the test directions have been submitted to you for: (1) review of both subject matter and adherence to measurement principles, and (2) suggestions for improvement.

The proposed instrument will assess teachers' knowledge of a subskills approach to teaching reading. The intended population is elementary teachers in grades one through six.

Directions:

1. Read the definitions of the three categories that comprise the test base.

2. Then review the items for the first category, one by one, and record your comments immediately following each item.

3. Repeat this procedure for the items in Categories II and III.

4. Finally, review the test directions.

While you are reading the items to judge if they are keyed to the appropriate category and that the answer is the best answer, you will also be on the lookout for measurement problems. Please use the following criteria as a guide to judge whether or not an item adheres to measurement principles:

1. Communicates clearly and explicitly

   -- free from ambiguity
   -- free from vagueness
   -- simple, direct wording
   -- words have precise meanings
   -- nonfunctional words excluded
   -- includes all qualifiers needed

2. Avoids clues to correct response

   -- pat verbal associations/sterotyped phrases
   -- grammatical construction
   -- common elements in stem and response
   -- correct answer consistently more precise and lengthy
   -- specific determiners (e.g., all, none, never)

3. General format is appropriate

   -- stem includes any words which could otherwise be repeated in the distractors
   -- distractors are parallel in grammatical structure
   -- item is editorially coherent and consistent

Evaluation Form for Each Item

1. Is the item keyed to the appropriate category? Yes ___ No ___

   If the keying isn't the best, why not? _____

   _____

   Which category would it fit in? _____

2. Do you agree that the answer is the best answer? Yes ___ No ___

   If not, please suggest a possible answer. _____

   _____

3. Does the item adhere to measurement principles? Yes ___ No ___

   If not, please suggest improvements. _____

   _____

4. General comments or suggestions not previously noted: _____

   _____

DIRECTIONS TO EXAMINEES

DIRECTIONS: Read each item carefully and decide which of the alternatives (answers) given is the best answer. Mark your answer to each item in the appropriate space on the answer sheet. Use a lead pencil to make a heavy, dark mark in the space that corresponds to the answer you select. Mark only one response per item. Since your answer sheet will be machine scored, erase completely any stray marks or answers that you wish changed so your response to that item will be counted.

Please answer every question, even though you may not be sure which alternative is the best answer.

Are the directions clear and concise? Yes _____ No _____

If not, please suggest improvements. _____

_____

_____

APPENDIX D

ITEMS FOR EACH CONTENT AREA IN THE PRELIMINARY TEST FORM

Items for Each Content Area in the Preliminary Test Form

CONTENT	TOPICS AND ITEMS
CATEGORY I: Identifying skills to be taught/learned (28 items)	Teachers know that the first step in organizing instruction is to decide what skills are most important for success in reading (item 1). Teachers know essential skills (items 2, 7, 8, 12, 13, 14, 16, 27, 29, 33, 41, 42, 44, 48, 54, 57, 61, 64, 65, 67, 70, 74). Teachers know about behavioral objectives (items 15, 21, 51, 66, 76).
CATEGORY II: Promoting skill development (42 items)	Teachers know that skills are taught routinely, at the student's rate of development (items 73, 86). Teachers know that skill instruction is usually provided to short-term groups with similar needs (item 34). Teachers know why records of skill development are kept (item 11). Teachers know about appropriate instruction of specific skills (items 3, 6, 18, 22, 24, 26, 30, 32, 35, 43, 45, 46, 47, 49, 52, 55, 56, 59, 60, 63, 69, 81, 83, 85). Teachers know about practice (items 58, 62, 71, 79, 80). Teachers know about transfer from skill lessons to reading texts (items 4, 9, 17, 20, 37, 40, 75, 77, 82).
CATEGORY III: Determining skill development status (16 items)	Teachers know how to assess skill development (items 19, 23, 25, 31, 36, 38, 50, 68, 72, 84). Teachers know about criterion-referenced testing (items 5, 10, 28, 39, 53, 78).

APPENDIX E

TEST FORM USED FOR FIELD TRYOUT I

(A) INFORMATION SHEET

(B) TEST

(C) ANSWER KEY

Please supply the following information.

1. How many reading courses have you taken? _____

2. How many hours of reading coursework have you taken?

   Semester hours _____ Quarter hours _____

3. At what grade level have you done most of your teaching?

   Primary _____ Intermediate _____

4. Indicate the number of years you have been a

   classroom teacher (grades one through six) _____
   reading teacher _____
   reading specialist _____
   other (please specify position/years) _____

5. Do you use (or have you used) a skills management system, e.g., Wisconsin Design for Reading Skill Development?

   Yes _____ No _____ If so, how many years _____

6. What is the school organization at your teaching level?

   self-contained _____ multi-unit _____ team _____

   other, please specify _____

NOTE: This information, as well as your test score, is confidential and will NOT be identified with you or your school at any time. It will ONLY be used to analyze the test results and to better understand the data.

DIRECTIONS: Read each item carefully and decide which on of the four alternatives, a, b, c, or d is the best answer. Mark your answer to each item in the appropriate space on the answer sheet. Use a soft lead pencil to make a heavy, dark mark in the space that corresponds to the answer you select. Mark only one response per item. Since your answer sheet will be machine scored, erase completely any stray marks to answers that you change, so your response to all items will be counted.

Please answer every item even though you may not be sure which alternative is the best answer.

NOTE: The term skill-centered reading program as used in this test is characterized as a reading program which provides for systematic instruction in the basic reading skills. The primary aim is to systematically teach skills to students. This instruction can be provided through the use of a basal reading series or an individualized or a language experience approach.

1. When the reading consultant and the teachers in a school develop a skill-centered reading program, the first step is to

   a. identify the skills to be taught.
   b. organize instructional materials.
   c. assess individual students' skill needs.
   d. place students in instructional groups.

2. Phonics instruction primarily helps students to

   a. identify irregular words.
   b. recognize meaningful word parts.
   c. pronounce unfamiliar regular words.
   d. spell words correctly in written work.

3. When students learn word identification skills, they should be encouraged to

   a. rely most on phonic analysis.
   b. rely on only one of the skills.
   c. work out unfamiliar words letter-by-letter.
   d. apply the skills both together and separately.

4. The goal of reading instruction is for students to use the skills they have learned to

   a. pronounce words correctly.
   b. read well both in and out of school.
   c. complete reading assignments correctly.
   d. pass tests administered after skills are taught.

5. Which statement would best describe John's performance on a criterion-referenced test?

   a. John has a percentile rank of 80.
   b. John is at the fifth-to-sixth-grade level.
   c. John got 70% of the items correct.
   d. John did better than 90% of the students.

6. The activity that is <u>most</u> useful for helping students acquire structural analysis ability is

   a. learning syllabication rules.
   b. building words with prefixes and suffixes.
   c. looking for little words in the big words.
   d. learning lists of prefixes and suffixes.

7. When students use other words in a sentence to figure out an unfamiliar word, they are using

   a. phonic analysis.
   b. visual analysis.
   c. contextual analysis.
   d. structural analysis.

8. When students identify topic sentences, determine sequence, and note details explicitly stated in the text, they have used

   a. literal comprehension skills.
   b. critical comprehension skills.
   c. creative comprehension skills.
   d. inferential comprehension skills.

9. Once students learn a word attack skill in developmental lessons, applying the skill in other contexts

   a. probably requires specific instruction.
   b. can be done independently by students.
   c. requires a minimum of teacher assistance.
   d. should not be scheduled, rather students should move on to another skill.

10. Performance standards that are established for criterion-referenced tests are usually

   a. set at different levels for individual students.
   b. set at a level which should be attained by students.
   c. invariant for word attack skills but flexible for other skills.
   d. scaled at 60%, 80%, and 100% for slow, average, and above average students.

11. The main reason for keeping records of individual students' skill development is to

   a. help teachers organize instruction.
   b. help teachers assign grades in reading.
   c. provide an incentive to learn more skills.
   d. monitor students' progress in their basal readers.

12. Basic sight words are those words students learn to recognize primarily through

   a. configuration.
   b. phonic analysis.
   c. contextual analysis.
   d. visual and auditory memory.

13. In comprehending a sentence, students need to

   a. attend to word order.
   b. use all the graphic features.
   c. pronounce all the words correctly.
   d. understand words and their grammatical relations.

14. The lists of reading skills such as those put together in skills management systems and basal reading programs

    a. have about the same number of skills.
    b. break skills into identical subskills.
    c. vary widely in terms of the number of skills.
    d. identify the same major categories of skills.

15. When a descriptive objective is written for a reading skill, a teacher specifies the

    a. tasks or learning activities students are to engage in.
    b. what students will learn as a result of instruction.
    c. frequency and quality of acceptable performance.
    d. level of competence students must attain.

16. When students show they understand a sentence by rearranging and/or substituting words without a subsequent change in meaning, they demonstrate the skill of

    a. paraphrase.
    b. contextual analysis.
    c. fluent oral reading.
    d. structural analysis.

17. When students learn a reading skill before other students in their instructional group, they should

    a. keep busy with an activity of their choice.
    b. start working on a new skill immediately.
    c. apply the skill in a variety of activities.
    d. complete the remaining exercises with their group.

18. Students are able to comprehend materials more easily if they

    a. read orally with expression.
    b. use phonics to identify words.
    c. use all the visual clues on a page.
    d. identify words accurately and automatically.

5

19. After students are taught to add the endings ed, ing, and s to familiar words and to pronounce the new words, the most appropriate test exercise will require students to

a. divide words into roots and endings.
b. pronounce words with ed, ing, and s endings that weren't used in instruction.
c. underline the ed, ing, and s endings in words.
d. pronounce a list of words used during instruction.

20. Application of skills that have been taught and learned is aided most when practice materials

a. are specially devised materials.
b. consist of various kits and games.
c. consist of workbook pages and ditto exercises.
d. closely resemble materials read in everyday life.

21. Standard behavioral objectives primarily specify

a. general educational goals.
b. what teachers will be doing.
c. what students are to achieve.
d. instructional procedures and tests.

22. Which of the following should not be used as a guideline for teaching reading comprehension skills?

a. Terminate instruction in a skill when students can pass a test.
b. Accept less sophisticated responses from younger students.
c. Adjust the content to match students' reading ability.
d. Model the comprehension process for students.

23. The paper-and-pencil exercise that is best for testing knowledge of symbol-sound correspondences is to have students

a. write the word the teacher says.
b. circle the word that begins with the same sound as the word the teacher says.
c. circle the word in which the underlined letter has the sound of the underlined letter in the nonsense word at the left.
d. circle the word in which the underlined letter has the sound of the underlined letter in the word at the left.

6

24. When students are unable to recognize easily confused letters (e.g., b-d, p-q), teachers should first

a. teach the letter names.
b. have students trace the letters.
c. emphasize the unique element in each pair.
d. give practice in discriminating geometric forms.

25. The best way to test recall of basic sight vocabulary is to ask students to

a. match pictures with the words.
b. read individual words on flashcards.
c. choose synonyms for target sight words.
d. choose target sight words for cloze blanks.

26. When basic reading skills are taught in developmental lessons, it is best to use

a. one instructional approach
b. one set of instructional materials.
c. a variety of materials and approaches.
d. a basal reading series with a strong phonics emphasis.

27. Two-letter consonant clusters which represent a sound unlike the sound of either individual letter (e.g., ng in gong or ch in reach) are

a. morphemes.
b. consonant blends.
c. consonant digraphs.
d. inflectional endings.

28. Teachers use criterion-referenced tests when they want to judge whether or not students have achieved at

a. grade level.
b. their ability level.
c. above or below average.
d. a desired level of proficiency.

29. Understanding the function of prepositions and conjunctions in sentences is most important for

a. identifying the main idea.
b. reading orally with expression.
c. noting relationships among ideas.
d. combining short sentences into longer ones.

30. Which completion exercise is most difficult for students?

a. The red ____ is in the vase.
b. The red r_s_ is in the vase.
c. The red r___ is in the vase.
d. The red r__e is in the vase.

31. Which of the following tasks is best for measuring the prereading skill of visual discrimination?

a. copying letters
b. matching letters
c. matching geometric shapes
d. writing letters from memory

32. When students can identify main ideas in fourth-grade basal reader stories, they

a. have mastered the main idea skill.
b. may need help in identifying main ideas in basal reader stories at other levels.
c. can identify main ideas in harder basal reader stories.
d. can also identify main ideas in a social studies text.

33. Which of the following would <u>not</u> be identified when using a structural analysis skill?

a. compound words
b. inflectional endings
c. prefixes and suffixes
d. consonant and vowel clusters

8

34. Skill-centered reading programs usually provide instruction in specific skills

a. to small groups of students with similar needs.
b. on an individual basis, as difficulties arise.
c. to the whole class at the end of the reading period.
d. to all students in a reading group, whenever the skills are introduced in the basal reader.

35. Instruction for developing reading comprehension skills is best characterized as helping students

a. progressively refine skills throughout the grades.
b. master different skills at each grade level.
c. learn skills in a hierarchical fashion.
d. master skills in a 1-2-3 order.

36. When a teacher wants to know whether or not students can use a certain skill, students should be asked to demonstrate

a. the skill in written material.
b. the prerequisites to the skill.
c. a simplified version of the skill.
d. the skill along with other skills.

37. The main purpose in providing time for recreational reading is to

a. get students to write book reports.
b. give students a change of pace once or twice a week.
c. let students use the skills learned in lessons.
d. give a teacher time to work with slow learners while others read.

38. In a skill-centered reading program, testing is usually done

a. after instruction in a skill.
b. at the end of the school year.
c. before instruction in a skill.
d. prior to and after teaching a skill.

39. The test which is most useful for assessing specific skill development is

a. a criterion-referenced test.
b. an informal reading inventory.
c. a standardized achievement test.
d. a standardized diagnostic reading test.

40. After students have learned to identify the central thought in basal reader stories, identifying the central thought in passages from a social studies text will usually

a. be no more difficult than in the basal reader stories.
b. be able to be done independently by students.
c. require teacher direction to learn to apply the skill.
d. be less complicated than in narrative materials.

41. Clusters of two or three consonant letters whose sound is a composite of each individual letter's sound (e.g., gr in grin or sl in slot) are

a. morphemes.
b. consonant blends.
c. consonant digraphs.
d. consonant diphthongs.

42. Dividing reading comprehension into main idea, supporting details, and sequencing skills is intended to

a. make teaching reading comprehension easier.
b. identify a simple to complex ordering of skills.
c. indicate students process main ideas before the other skills.
d. imply main ideas should be mastered first, then details, etc.

43. In general, reading authorities agree that phonics instruction should begin with

a. long vowel sounds.
b. word families.
c. short vowel sounds.
d. single consonant sounds.

44. The most important criterion for selecting sight words is that the words

a. are troublesome to learn.
b. can be decoded with phonics.
c. are part of students' speaking vocabularies.
d. are common in printed materials.

45. The activity that is <u>least</u> useful for helping students acquire structural analysis ability is

a. relating letters to sounds.
b. identifying meaningful word parts.
c. discussing the meaning of the words.
d. adding prefixes and suffixes to words.

46. When a test given after developmental instruction and practice indicates that a skill has not been learned, a teacher should

a. let the skill develop on its own.
b. reteach the skill using alternative techniques.
c. use inductive rather than deductive teaching techniques.
d. reteach the skill with the same materials and procedures.

47. A number of techniques have proved successful for teaching letter-sound associations. The technique for which stronger support exists is teaching

a. letter-by-letter sounding.
b. letter sounds in the context of words.
c. letter sounds first, then blending them into words.
d. irregular sounds of a letter when it is introduced.

48. Research literature on reading comprehension indicates that comprehension is

a. one general skill.
b. several basic skills.
c. nine separate skills.
d. 80 or more supposedly discrete skills.

49. To make the best use of prefixes and suffixes when identifying unfamiliar words, students should

    a. apply several meanings of a prefix or suffix to a root word.
    b. learn the meanings of a large number of prefixes and suffixes.
    c. assign meaning to a prefix or suffix before identifying the root.
    d. decode prefixes and suffixes after phonics is applied to the word.

50. In skill-centered reading programs, teachers determine students' skill needs with

    a. individually administered tests.
    b. standardized reading survey tests.
    c. oral reading performance in the basal reader.
    d. brief, group and/or individually administered tests.

51. Which of the following statements is a behavioral objective?

    a. The student will be able to correctly alphabetize a list of ten words.
    b. An analysis of Barrett's Taxonomy will be presented within a 30 minute period.
    c. Following a discussion of blends, students will understand what they are.
    d. The teacher will explain the importance of digraphs to the third grade class.

52. When a teacher wishes to develop ability in inferential reading, instruction should be provided

    a. after students reach fourth grade.
    b. after students can read grade two materials.
    c. after literal comprehension is completely mastered.
    d. at the beginning reading levels and on through successive grades.

53. The major difference between criterion-referenced tests and norm-referenced tests is the

    a. content of the test.
    b. meaning of the scores.
    c. format of the test items.
    d. way comprehension is tested.

54. Students can figure out the meaning of an unfamiliar word in a sentence best by using

a. phonics.
b. word parts.
c. word parts and context.
d. context.

55. When students are learning to identify the central thought of a passage, they usually find it easiest to

a. generate their own central thought for the passage.
b. identify a central thought statement in the passage.
c. distinguish between relevant and irrelevant details.
d. identify the general topic for a passage from a given list.

56. The first few steps in a skill lesson will promote learning of the skill for most students if a teacher

a. uses programmed instructional materials.
b. patterns the steps as a basal reading lesson does.
c. demonstrates the skill and has students practice working with the skill.
d. asks students to complete worksheets independently.

57. Using contextual analysis to identify an unfamiliar word helps students determine the word's

a. intended meaning.
b. meaningful parts.
c. precise definition.
d. exact pronounciation.

58. Practice of skills that have been taught is aided most when students

a. receive feedback on correct and incorrect answers.
b. review the materials that were used in instruction.
c. do a large number of worksheets during each session.
d. are directly assisted through each step of an activity.

59. Most reading authorities recommend that reading comprehension skills should be learned and practiced in

a. content area materials as part of the regular assignment.
b. guided activities usually with basal reader materials.
c. specially devised materials until mastery on a test.
d. independent activities with comprehension kits.

60. Which of the following activities for students is least appropriate for learning to recognize basic sight words?

a. Reading stories that contain the words.
b. Relating letters to sounds in the words.
c. Repeatedly tracing and saying the words.
d. Noting the visual characteristics of the words.

61. Which one of the following statements does not describe inflectional endings?

a. Inflectional endings are morphemes.
b. Inflectional endings have invariant meanings.
c. Inflectional endings may change tense or number of a word.
d. Inflectional endings may change the part of speech of a word.

62. After a lesson on pronouncing printed words with the at spelling pattern has been taught, the most appropriate practice exercise would require students to

a. make up new at words.
b. pronounce at words in print.
c. underline at words in sets of words.
d. discriminate at words in spoken words.

63. One of the most important factors in teaching reading comprehension skills is the

a. questions teachers ask in discussions.
b. use of kits, workbooks, and A-V materials.
c. sequence in which the skills are taught/learned.
d. number of worksheets students complete independently.

64. The word identification skill that is used most frequently by competent readers is

a. syllabication.
b. phonic analysis.
c. contextual analysis.
d. structural analysis.

65. The reading comprehension skill that is most basic for understanding a selection is

a. grasping the main idea.
b. recognizing relevant details.
c. knowing the meanings of words.
d. restating the selection in one's own words.

66. Once the skills are identified for a reading program, they are put together in a list. This list should

a. be a random ordering of skills.
b. tell the grade levels to teach the skills.
c. show logical sequences for teaching the skills.
d. follow the ordering in well known basal readers.

67. When students use structural analysis, an unfamiliar word is worked out

a. on a letter-by-letter basis.
b. by examining its meaningful parts.
c. by the letter-sound correspondences.
d. through the structure of the sentence.

68. Before teaching the skill of sound blending, a teacher should first assess students'

a. letter-name knowledge.
b. letter-sound knowledge.
c. level of perceptual motor skills.
d. reading readiness with standardized tests.

69. If several students in a class do not know the digraph skills, the best way to develop competence in these skills is to

a. provide direct assistance in using the skills.
b. let the skills develop in wide reading.
c. emphasize the skills in the workbook.
d. use basal reading lessons.

70. While searching for a main idea in a selection, students will need to use

a. convergent thinking skills.
b. comprehension skills of analysis.
c. comprehension skills of synthesis.
d. a combination of analysis, synthesis, and convergent thinking skills.

71. Practice exercises generally produce better results when they are

a. spaced over days or weeks during and after instruction.
b. done immediately after instruction, as a set of activities.
c. provided in the form of massed practice, after instruction.
d. distributed over the period in which instruction is being given.

72. One of the best indicators of reading comprehension is the

a. score on a standardized reading test.
b. ability to read fluently with expression.
c. ability to paraphrase or recognize paraphrase.
d. score on a timed reading comprehension test.

73. In a skill-based reading program, the usual procedure is to teach reading skills

a. when students ask for help.
b. when individual difficulties arise.
c. regularly as part of the daily schedule.
d. when students lag in their skill development.

74. When students use syntactic clues to figure out unfamiliar words, they rely on

a. structural analysis.
b. meaningful word parts.
c. syllabication patterns.
d. word order in sentences.

75. After students have learned a skill in developmental instruction, a teacher should

a. present materials in which students can apply the skill.
b. reteach the skill when it comes up in the basal reader.
c. provide review activities just like the instructional ones.
d. consider the skill mastered if worksheets are done correctly.

76. When the skills that will be taught in a skill-centered reading program have been identified, the next step is to

a. write behavioral objectives.
b. specify relevant learning activities.
c. construct criterion-referenced tests.
d. write behavioral and descriptive objectives.

77. After students learn a skill and then practice it in a variety of tasks rather than the same type of tasks, they will usually

a. forget the skill sooner.
b. use it whenever they read.
c. require fewer practice sessions.
d. require longer practice sessions.

78. A skill group made up of students in grades two to five was given a criterion-referenced test on a word attack skill. The teacher should decide who passed the test by

a. determining passing scores on an individual basis.
b. using a lower criterion for primary grade students.
c. determining which students scored at or above the established criterion.
d. scaling criterion levels for slow, average, and above average students.

79. Reading authorities suggest that the best way to insure that students learn and remember basic sight words is to

a. practice with flash-word cards.
b. practice with flash-phrase cards.
c. use phonics to sound out the words.
d. let the words develop on their own.

80. When students show understanding of a skill in a skill lesson and then later forget it, this probably can be traced to

a. lack of ability.
b. lack of practice.
c. lack of modeling.
d. insufficient rewards.

81. After a comprehension skill (e.g., determining the order of events) has been taught and students pass a test on the skill, the skill

a. is considered mastered, once and for all.
b. probably still needs further refinement.
c. needs little or no further developmental instruction.
d. can be applied independently in varying contexts and levels.

82. Students will be better able to remember and apply a skill if they practice the skill in materials that

a. were used in instruction.
b. are unlike those used in instruction.
c. are just like those used in instruction.
d. are different but still related to the instructional ones.

83. When students are unable to pronounce words which contain an oo combination after instruction, they will probably learn the skill if it is

a. repeated with the same instruction.
b. taught along with related skills.
c. retaught with different materials.
d. left to develop in independent reading.

84. After skills that teachers intend to teach are described in terms of behavioral objectives, teachers

a. begin teaching the skills right away.
b. identify individual students' skill needs.
c. break the skills down into very specific subskills.
d. decide which skills should be dropped from the list.

85. When cloze exercises are used as instructional activities, the words students supply for those deleted should be

a. spelled correctly.
b. similar in meaning.
c. similar phonetically.
d. similar syntactically.

86. The rate at which skills are introduced and taught in a skill-centered reading program should be determined by

a. a student's grade level.
b. the basal reading series.
c. a student's general achievement level.
d. the continual reassessment of student progress.

ANSWER KEY

1. a	23. c	45. a	66. c
2. c	24. c	46. b	67. b
3. d	25. b	47. b	68. b
4. b	26. c	48. b	69. a
5. c	27. c	49. a	70. d
6. b	28. d	50. d	71. a
7. c	29. c	51. a	72. c
8. a	30. a	52. d	73. c
9. a	31. b	53. b	74. d
10. b	32. b	54. c	75. a
11. a	33. d	55. d	76. d
12. d	34. a	56. c	77. b
13. d	35. a	57. a	78. c
14. c	36. a	58. a	79. b
15. a	37. c	59. a	80. b
16. a	38. d	60. b	81. b
17. c	39. a	61. d	82. d
18. d	40. c	62. b	83. c
19. b	41. b	63. a	84. b
20. d	42. a	64. c	85. b
21. c	43. d	65. c	86. d
22. a	44. d		

APPENDIX F

INSTRUCTIONS FOR EXAMINERS - FIELD TRYOUT I

General Instructions

It is important that you are familiar with the PROCEDURES you are to follow as well as the DIRECTIONS to the examinees before the testing session begins.

Please read the instruction to the examinees slowly and clearly in a voice that is loud enough to be heard. Hold faithfully to the wording of the printed directions.

Teachers should not receive telephone calls or visitors during the testing sessions. Please take messages for them.

Specific Instructions

1. BEFORE testing

   a. Check the noise, lighting, and physical features of the appointed room. Relocate for testing if environmental factors could adversely affect the test results.

   b. Have the tests, answer sheets, and extra pencils ready for distribution.

2. DURING the session

   (Distribute the tests and the answer sheets after the examinees are assembled.)

   Examiner (read):

   Please fill out the information sheet. When everyone is done we'll go on to the next page. This information is confidential and will only be used in discussing the test as a whole. Individual scores cannot be identified with persons or schools.

   (After the information requested is recorded.)

   Examiner (read):

   Open your test booklet to the next page. Please follow along as I read the DIRECTIONS:

   Read each item carefully and decide which one of the four alternatives, a, b, c, or d is the best answer. Mark your answer to each item in the appropriate space on the answer sheet.

Specific Instructions
Page 2

Use a soft lead pencil to make a heavy, dark mark in the space that corresponds to the answer you select. Mark only one response per item. Since your answer sheet will be machine scored, erase completely any stray marks to answers that you change, so your response to all items will be counted.

Please answer every item even though you may not be sure which alternative is the best answer.

NOTE: The term skill-centered reading program as used in this test is characterized as a reading program which provides for systematic instruction in the basic reading skills. The primary aim is to systematically teach skills to students. This instruction can be provided through the use of a basal reading series or an individualized or language experience approach.

Does anyone have any questions about the directions?

(Answer questions, if any, but do NOT try to further define skill-centered reading program.)

I cannot answer questions you may have about individual items Just make a sincere effort to do your best. The test is not timed. It should take about 45 minutes. After you are done, please return the test booklet and your answer sheet to me.

3. AFTER testing

Promptly return the test booklets and answer sheets in a manilla envelope to the office.

APPENDIX G

ITEMS FOR EACH CONTENT AREA IN THE REVISED TEST FORM

Items for Each Content Area in the Revised Test Form

CONTENT	TOPICS AND ITEMS
CATEGORY I: Identifying skills to be taught/learned (16 items)	Teachers know that the first step in organizing instruction is to decide what skills are most important for success in reading (item 34).
	Teachers know essential skills (items 3, 4, 16, 21, 33, 36, 38, 39, 44, 46, 53, 56).
	Teachers know about behavioral objectives (items 9, 28, 41).
CATEGORY II: Promoting skill development (31 items)	Teachers know that skills are taught routinely at the student's rate of development (items 48, 60).
	Teachers know that skill instruction is usually provided to short-term groups with similar needs (item 22).
	Teachers know why records of skill development are kept (item 17).
	Teachers know about appropriate instruction of specific skills (items 2, 5, 6, 7, 11, 13, 15, 20, 24, 30, 32, 37, 42, 47, 52, 57, 58).
	Teachers know about practice (items 8, 18, 31, 45, 49, 51, 59).
	Teachers know about transfer from skill lessons to reading texts (items 10, 14, 26).
CATEGORY III: Determining skill development status (13 items)	Teachers know how to assess skill development (items 12, 23, 27, 35, 40, 43, 55).
	Teachers know about criterion-referenced testing (items 1, 19, 25, 29, 50, 54).

APPENDIX H

INSTRUCTIONS AND EVALUATION FORM

FOR REASSESSMENT OF CONTENT VALIDITY

Reviewer _____ Position _____

Date _____

To the reviewer: The purpose of this second review is to establish content validity for the test. You are to determine if the three categories of the test base are tested by the items.

The decision was to retain the set of items in a .31 to .91 difficulty range. Although a larger number of items came through for Category II, the categories will not be equalized (as called for by the test specifications) because to do so could lower reliability. Further, the category has more divisions and items about promoting skill development are important. Items have been revised (e.g., poor distractors, awkward wording, etc.) and it was necessary to write several new items (17 + 18) for Category I (B). Items are not included for portions of the test base [Category I (B) and Category II (B)] because differences of opinion among skill people do not provide a legitimate basis for constructing items. Hopefully, all opinion items have been eliminated. Two questions follow each item and are interpreted as designated.

1. Right Category? Yes ___ No ___ is interpreted as: Is this item keyed to the right category?

2. Best answer? Yes ___ No ___ is interpreted as: Is the designated answer the best answer?

   Please feel free to make suggestions, deletions, rewording, etc. on the items themselves.

Are the three categories that comprise the test base tested by the items? Yes ___ No ___

_____

_____

_____

_____

_____

APPENDIX I

TEST FORM USED FOR FIELD TRYOUT II

(A) INFORMATION SHEET

(B) TEST

(C) ANSWER KEY

Please supply the following information.

1. How many reading courses have you taken? _____

   (count courses equivalent to 2-3 semester credit hours)

2. How many years have you worked as a classroom teacher in grades one through six? _____

3. Check the grade level at which you have done most of your teaching.

   primary _____ ____ intermediate _____

4. Check if you use (or have used) a skills management system, e.g., Wisconsin Design for Reading Skill Development.

   Yes _____ No _____

READING TEST

DIRECTIONS: Read each item carefully and decide which one of the four alternatives, a, b, c, or d is the best answer. Use a soft lead pencil to make a heavy, dark mark in the space on the answer sheet that corresponds to the answer you select. Since your answer sheet will be machine scored, erase completely any stray marks to answers that you change, so your response to all items will be counted.

There is only one BEST answer for each item. Please answer all the items.

1. The test which is most useful for assessing specific skill development is

   a. a cloze test.
   b. a criterion-referenced test.
   c. a standardized survey reading test.
   d. an informal reading inventory published by the basal series.

2. When learning to use word identification skills, students should be encouraged to work out words by

   a. depending on context clues.
   b. relying most on phonic analysis.
   c. applying phonics before structural analysis.
   d. applying the skills both together and separately.

3. Two-letter consonant clusters which represent a sound unlike the sound of either individual letter (e.g., gh in cough or ch in reach) are

   a. morphemes.
   b. consonant blends.
   c. consonant digraphs.
   d. inflectional endings.

4. When students use structural analysis, an unfamiliar word is usually worked out

   a. through meaningful parts (e.g., affixes, root words).
   b. through letter clusters and spelling patterns.
   c. through the structure of the sentence.
   d. by the letter-sound correspondences.

5. When students are learning to identify the central thought of a passage, they usually find it easiest to

   a. state their own central thought for the passage.
   b. identify a central thought statement in the passage.
   c. distinguish between relevant and irrelevant details.
   d. identify the general topic for a passage from a given list.

6. The activity that is least appropriate for learning to recognize basic sight words is

   a. reading stories that contain the words.
   b. relating letters to sounds in the words.
   c. practicing writing and saying the words.
   d. noting the visual characteristics of the words.

7. In general, reading authorities suggest that phonics instruction should initially emphasize

   a. single consonant sounds.
   b. short vowel sounds.
   c. long vowel sounds.
   d. word families.

8. When students show understanding of a skill in a skill lesson and then later forget it, this probably can be traced to a lack of

   a. ability.
   b. practice.
   c. relevance.
   d. oral instruction.

9. Objectives that are written for word attack skills usually describe

   a. specific instructional activities.
   b. acceptable grade level performance.
   c. the expected outcome of instruction.
   d. learner behaviors during instruction.

10. After students have learned to identify the central thought in basal reader stories, identifying the central thought in passages from a social studies text will usually

    a. be about as difficult as in the basal reader.
    b. be able to be done independently by students.
    c. be less complicated than in narrative materials.
    d. require teacher assistence to learn to apply the skill.

11. The completion exercise that is most difficult for students is

    a. The red ____ is in the vase.
    b. The red r_s_ is in the vase.
    c. The red r___ is in the vase.
    d. The red r__e is in the vase.

12. When teachers want to determine whether or not students know a certain skill, students should be asked to demonstrate the

    a. subskills that make up the skill.
    b. skill in content used in instruction.
    c. skill in tasks similar to those used in instruction.
    d. skill at levels and in contexts not emphasized in instruction.

13. After a comprehension skill has been taught and students pass a test on it, the skill probably

    a. does not need to be retested.
    b. requires no further instruction because it is mastered.
    c. can be applied independently in varying contexts and levels.
    d. needs further instruction at other levels and in other contexts.

14. The main purpose for providing time for recreational reading is to

    a. increase poorer readers' rate of reading.
    b. let students use the skills learned in lessons.
    c. increase better readers' informational background.
    d. give students a change of pace once or twice a week.

15. The activity that is most helpful for developing structural analysis ability is

    a. learning syllabication rules
    b. building words with prefixes and suffixes.
    c. looking for little words in the big words.
    d. learning the meanings of many prefixes and suffixes.

4

16. Understanding the function of prepositions and conjunctions in sentences is most important for

    a. identifying details.
    b. reading orally in thought units.
    c. noting relationships among ideas.
    d. combining short sentences into longer ones.

17. The main reason for keeping records of the skills each student has and has not mastered is to

    a. communicate with parents.
    b. help teachers plan and organize instruction.
    c. compare a student's performance to the performance of others.
    d. insure that each student is taught the skills designated for his grade level.

18. Students will be better able to remember a skill if they practice the skill

    a. until a test indicates the skill is mastered.
    b. after instruction is given, with a set of worksheets.
    c. with varied activities during and immediately after instruction.
    d. during instruction and periodically (over weeks/months) after instruction.

19. Performance standards that are established for criterion-referenced tests are usually

    a. set at different levels for individual students.
    b. set at a level which should be attained by all students.
    c. set at one level for word attack skills but flexible for other skills.
    d. scaled at 60%, 80%, and 100% for slow, average, and above average students.

20. When cloze exercises are used for instruction, the words accepted as correct for those deleted would probably be

    a. exact replacements.
    b. reasonable synonyms.
    c. words of the same length.
    d. words of the same grammatical class.

21. A cluster of two or three consonant letters whose sound is a composite of the sounds of each of the individual letters (e.g., sl in slot or str in strong) is a

    a. consonant digraph.
    b. consonant blend.
    c. morpheme.
    d. phoneme.

22. If the focus of a reading program is teaching specific skills, instruction is most efficient and effective when it is provided

    a. on an individual basis, as difficulties arise.
    b. to small, short-term groups of students with similar needs.
    c. by teaching the skills designated for a particular grade level.
    d. to students in a reading group when the skills are introduced in the basal reader.

23. The best way to test recognition of basic sight vocabulary is to ask a student to

    a. match pictures with the words.
    b. read individual words on flashcards.
    c. choose sight words for close blanks.
    d. pronounce sight words when occurring in context.

24. The activity that is least useful for helping students develop structural analysis ability is

    a. relating letters to sounds.
    b. building and defining compound words.
    c. changing the number and tense of words.
    d. defining words composed of root words and affixes.

25. The major difference between criterion-referenced tests and norm-referenced tests is the

    a. content of the tests.
    b. meaning of the scores.
    c. format of the test items.
    d. way comprehension is tested.

26. Application of a skill that has been taught and learned is aided most when students practice the skill with

a. basal reader materials.
b. a set of worksheets and/or in a kit.
c. materials that are identical to those used in instruction.
d. varied reading tasks (e.g., content area, practical contexts).

27. If the focus of a reading program is teaching specific skills, testing is most useful if it is done

a. after instruction in a skill.
b. at the beginning and end of the school year.
c. after units in the basal reader are completed.
d. to determine needs and after a skill is taught.

28. In stating behavioral objectives, one primarily specifies

a. general educational goals.
b. what teachers will be doing.
c. what students are to achieve.
d. instructional procedures and tests.

29. A skill group made up of students in grades two to five was given a criterion-referenced test on a word attack skill. The teacher should decide who has mastered the skill by

a. determining which students scored at or above the established criterion.
b. determining passing scores on an individual basis.
c. using a lower criterion for primary grade students.
d. scaling criterion levels for slow, average, and above average students.

30. After students can identify main ideas in fourth-grade basal reader stories, they

a. have mastered the main idea skill.
b. can also identify main ideas in a social studies text.
c. can identify main ideas in other reading selections.
d. may need help in identifying main ideas in basal reader stories at other levels.

31. If students practice a skill in a variety of reading tasks rather than in the same type of task, they will usually

    a. tend to become confused.
    b. require fewer practice sessions.
    c. be more prone to forget the skill.
    d. use the skill more readily when they read.

32. Instruction for developing reading comprehension skills is best characterized as helping students

    a. learn skills in a sequential order.
    b. progressively refine skills throughout the grades.
    c. master a set of different skills at each grade level.
    d. master literal comprehenion before inferential comprehension.

33. When students identify topic sentences, determine sequence, and note details explicitly stated in the text, they have used

    a. literal comprehension skills.
    b. critical comprehension skills
    c. evaluative comprehension skills.
    d. inferential comprehension skills.

34. After the reading consultant and the teachers in a school have decided to develop a reading program which is focused on teaching skills, their first step would be to

    a. assess individual students' skill needs.
    b. designate skills to be taught at each grade level.
    c. identify the skills that are important for students to learn.
    d. key materials and procedures to the objectives of the program.

35. When the focus of a reading program is teaching skills, students' specific skill needs are best identified with

    a. standardized survey reading tests.
    b. workbook tests accompanying the basal reader.
    c. daily worksheets and informal teacher observation.
    d. brief, group and/or individually administered tests.

8

36. When students use syntactic clues to figure out unfamiliar words, they primarily rely on

a. structural analysis.
b. meaningful word parts.
c. word order in sentences.
d. letter clusters and spelling patterns.

37. When students are unable to recognize easily confused letters (e.g., b-d, p-q), teachers should initially

a. teach the letter names.
b. have students trace the letters.
c. emphasize the unique element in each pair.
d. give practice in matching geometric forms.

38. Students can best figure out the meaning of an unfamiliar word in a sentence by using

a. structural analysis.
b. clues within the word.
c. all the graphic features.
d. word parts and the context.

39. When students show they understand a sentence by rearranging and/or substituting words without a subsequent change in meaning, they demonstrate the skill of

a. paraphrase.
b. contextual analysis.
c. structural analysis.
d. interpretive reading.

40. The task that is most appropriate for assessing a prereading visual discrimination skill is

a. reproducing geometric forms.
b. matching letters and words.
c. writing letters from memory.
d. copying letters and patterns.

41. Objectives that are written for skills would least likely be used to

a. work out instructional sequences.
b. develop tests to assess student progress.
c. identify instructional materials and procedures.
d. determine if students are performing at grade level.

42. When basic reading skills are taught in skill lessons, it is best to use

a. workbook-type activities and games.
b. one instructional method or approach.
c. a variety of materials and approaches.
d. basal reading materials with a strong decoding program.

43. After students are taught to add the endings ed, ing, and s to familiar words and to pronounce the new words, the most appropriate test exercise is

a. dividing words into roots and endings.
b. underlining ed, ing, and s words in sets of words.
c. making up new words by adding ed, ing, and s to root words.
d. pronouncing words with ed, ing, and s endings that weren't used in instruction.

44. When students use context clues to identify an unfamiliar word, they determine the word's

a. intended meaning.
b. oral counterpart.
c. structural units.
d. precise definition.

45. Practice of skills that have been taught is aided most when students

a. do more written than oral activities.
b. receive feedback on correct and incorrect answers.
c. repeat the materials that were used in instruction.
d. are directly assisted through each step of an activity.

46. When using a structural analysis skill, students would not usually identify

a. compound words.
b. plural and tense endings.
c. consonant or vowel clusters.
d. contractions and possessives.

47. When a teacher wishes to develop ability in inferential reading, instruction should be provided

a. after students learn the basic word attack skills.
b. after literal comprehension is completely mastered.
c. after students have learned to read orally with some fluency.
d. at the beginning reading levels and on through successive grades.

48. The rate at which skills are introduced and taught in a skill-centered reading program should be determined by

a. a student's general achievement level.
b. the instructional materials available.
c. a continual reassessment of student progress.
d. the number of skills designated for each grade level.

49. Reading authorities generally recommend that reading comprehension skills should be learned and practiced primarily with

a. short, high-interest narrative selections.
b. content area materials as subject matter is learned.
c. guided activities usually using basal reader materials.
d. specifically devised materials directed toward mastery on a test.

50. The statement that best describes John's performance on a criterion-referenced test is that John

a. got 70% of the items correct.
b. has a percentile rank of 80.
c. did better than 90% of the students.
d. is at the fifth-to-sixth grade level.

51. Students are able to comprehend materials more easily if they

a. read the materials orally.
b. rely on phonics to decode words
c. use all the visual clues on a page.
d. identify words accurately and automatically.

52. When students are unable to pronounce words which contain an oo combination after instruction, they will most likely learn the skill if it is

a. retaught with different materials
b. repeated with the same instruction.
c. taught along with other vowel digraphs.
d. retaught when clusters with similar sounds are taught.

53. The main purpose for phonics instruction is to help students

a. decode irregular words.
b. recognize meaningful word parts.
c. pronounce unfamiliar regular words.
d. identify each of the phonemes within a word.

54. Teachers use criterion-referenced tests when they want to judge whether or not students have achieved at

a. grade level.
b. their ability level.
c. the 80th percentile.
d. a desired level of proficiency.

55. One of the best indicators of reading comprehension is the

a. ability to read sentences fluently.
b. score on a timed reading comprehension test.
c. ability to paraphrase or to recognize a paraphrase.
d. score on a standardized test of reading vocabulary.

56. The word identification skill that is probably used most frequently by competent readers is

a. contextual analysis.
b. structural analysis.
c. phonic analysis.
d. syllabication.

57. One of the most important factors in teaching reading comprehension skills is the

a. questions teachers ask in discussions of what is read.
b. type of worksheets students complete independently.
c. sequence in which the skills are taught/learned.
d. questions students answer in the workbook.

58. A number of techniques have proved successful for teaching letter-sound associations. The technique for which stronger support exists is teaching

a. variable sounds of a letter within the same lesson.
b. letter sounds within the context of familiar words.
c. letter-by-letter sounding with words of one syllable.
d. certain letter sounds which are later blended into words.

59. Reading authorities generally suggest that the best way to insure that students learn and remember basic sight words is to

a. analyze the words with phonics.
b. drill the words with flash cards.
c. practice the words in phrases and sentences.
d. recognize the words by their distinctive visual forms.

60. In a skill-centered reading program, the usual procedure is to teach specific skills

a. when individual difficulties arise.
b. as sequenced in the basal reading series.
c. as needed in the reading of specific stories.
d. regularly, moving each student along at his/her own rate.

ANSWER KEY

1. b	21. b	41. d
2. d	22. b	42. c
3. c	23. b	43. d
4. a	24. a	44. a
5. d	25. b	45. b
6. b	26. d	46. c
7. a	27. d	47. d
8. b	28. c	48. c
9. c	29. a	49. b
10. d	30. d	50. a
11. a	31. d	51. d
12. c	32. b	52. a
13. d	33. a	53. c
14. b	34. c	54. d
15. b	35. d	55. c
16. c	36. c	56. a
17. b	37. c	57. a
18. d	38. d	58. b
19. b	39. a	59. c
20. b	40. b	60. d

APPENDIX J

INSTRUCTIONS FOR EXAMINERS - FIELD TRYOUT II

General Instructions

It is important that you are familiar with the PROCEDURES you are to follow as well as the DIRECTIONS to the examinees before the testing session begins.

Please follow the instructions precisely in administering the test. In so doing, the score variance that is due to factors external to the examinees will be minimized.

Please read the instructions to the examinees slowly and clearly in a voice that is loud enough to be heard. Hold faithfully to the wording of the printed directions.

Teachers should not receive telephone calls or visitors during the testing session. Please take messages for them.

Specific Instructions

1. BEFORE testing

   a. Check the noise, lighting, and physical features of the appointed room. Relocate for testing if the environmental factors could adversely affect the test results.

   b. Have the test, answer sheets, and extra pencils ready for distribution.

2. DURING the session

   (Distribute the tests and the answer sheets after the examinees are assembled.)

   Examiner (read):

   Please fill out the information sheet. When everyone is done we'll go on to the next page. This information is confidential and will only be used in discussing the test as a whole. Individual scores cannot be identified with persons or schools

   (After the information requested is recorded.)

Specific Instructions
Page 2

Examiner (read):

Open your test booklet to the next page. Please follow along as I read the DIRECTIONS:

Read each item carefully and decide which one of the four alternatives, a, b, c, or d is the best answer. Use a soft lead pencil to make a heavy, dark mark in the space on the answer sheet that corresponds to the answer you select. Since your answer sheet will be machine scored, erase completely any stray marks to answers that you change, so your response to all items will be counted.

There is only one BEST answer for each item. Please answer all the items.

Does anyone have any questions about the directions?

(Answer questions, if any)

I cannot answer questions you may have about individual items. Just make a sincere effort to do your best. The test is not timed. It should take about 40 minutes. After you are done, please return the test booklet and your answer sheet to me.

3. AFTER testing

Promptly return the test booklets and answer sheets in a manilla envelope to the office.